

**Metaanalysen zum Zusammenhang von der Allgemeinen,  
Fluiden und Kristallisierten Intelligenz  
mit dem Fünf-Faktoren-Modell der Persönlichkeit unter  
besonderer Berücksichtigung des Publikationsbias**

Inaugural Dissertation  
zur Erlangung der Doktorwürde  
der Philosophischen Fakultät  
der Rheinischen Friedrich-Wilhelms-Universität zu Bonn

vorgelegt von  
**Sebastian Lange**

aus Sondershausen

Bonn, 2013

**Gedruckt mit der Genehmigung der Philosophischen Fakultät der  
Rheinischen Friedrich-Wilhelms-Universität Bonn**

**Zusammensetzung der Prüfungskommission:**

Vorsitzender: PD Dr. Bernd Schlöder

Betreuer: Prof. Dr. André Beauducel

Gutachter: Prof. Dr. Gerhard Blickle

Weiteres prüfungsberechtigtes Mitglied: Jun.-Prof. Dr. Anja Leue

Tag der mündlichen Prüfung: 05.06.2013

*“Ja, wir werden alles, alles noch einmal in Frage stellen. Und wir werden nicht mit Siebenmeilenstiefeln vorwärtsgehen, sondern im Schnecken tempo. Und was wir heute finden, werden wir morgen von der Tafel streichen und erst wieder anschreiben, wenn wir es noch einmal gefunden haben. Und was wir zu finden wünschen, das werden wir, gefunden, mit besonderem Mißtrauen ansehen.“ (...)* „Sollte uns dann aber jede andere Annahme als diese unter den Händen zerronnen sein, dann keine Gnade mehr mit denen, die nicht geforscht haben und doch reden.“

*(Brecht, 1963, S.93).*

## **Abstract**

In der vorliegenden Arbeit wurden metaanalytische Untersuchungen zum Zusammenhang der Fähigkeitskonstrukte Allgemeine Intelligenz (g), Fluide Intelligenz (Gf) und Kristallisierten Intelligenz (Gc) mit den Persönlichkeitsdimensionen Neurotizismus, Extraversion, Offenheit für Erfahrung, Verträglichkeit und Gewissenhaftigkeit durchgeführt. Da von der Einschätzung ausgegangen wurde, dass die Autoren der publizierten Studien eine z.T. liberale bzw. ungenaue Zuordnung der jeweiligen Intelligenztests bzw. kognitiven Leistungsaufgaben zu den Fähigkeitskonstrukten g, Gf oder Gc vorgenommen haben, wurden zwei Analyse-Varianten gerechnet. In der ersten Analyse-Variante wurde die Klassifizierung der Fähigkeitskonstrukte, wie sie von den Autoren der publizierten Studien erfolgte, beibehalten. Dadurch ist die Variabilität in der Zuordnung von Aufgaben zu Fähigkeitskonstrukten in der Variabilität der Effektstärken enthalten. In einer anschließenden zweiten Analyse-Variante wurde eine nachträgliche Zuordnung der Fähigkeitskonstrukte gemäß der Cattell-Horn-Carroll-Theorie (McGrew, 1997, 2005, 2009) nach der Zuordnungsvorlage von Süß und Beauducel (2011) durch den Autor der vorliegenden Arbeit vorgenommen. Ein Vergleich der beiden Analysen erlaubt eine Einschätzung der Auswirkungen unterschiedlicher Zuordnungen der Fähigkeitskonstrukte auf die metaanalytischen Kennwerte. Die Ergebnisse zeigen z.T. deutliche Unterschiede in Abhängigkeit der Analyse-Variante. Insgesamt sind die untersuchten Populationszusammenhänge als klein einzuschätzen. Unabhängig von den z.T. unterschiedlichen Ergebnissen in Abhängigkeit des verwendeten Verfahrens zur Identifikation eines Publikationsbias ist davon auszugehen, dass in mindestens 50% der untersuchten Zusammenhänge ein Publikationsbias auftritt.

## Inhalt

|  |            |
|--|------------|
| <b>Danksagung .....</b>  | <b>III</b> |
| <b>1 Einleitung.....</b>   | <b>1</b>   |
| <b>2 Theorie .....</b>   | <b>4</b>   |
| 2.1 Für die vorliegende Arbeit relevante Intelligenzstrukturmodelle .....  | 4          |
| 2.2 Fünf-Faktoren-Modell der Persönlichkeit .....  | 8          |
| 2.3 Zusammenhang von Intelligenz und Persönlichkeit.....   | 11         |
| 2.3.1 Gemeinsamkeiten und Unterschiede von Intelligenz und<br>Persönlichkeit .....   | 12         |
| 2.3.2 Modelle zum Zusammenwirken von Intelligenz und Persönlichkeit .....  | 15         |
| 2.3.3 Metaanalytische Ergebnisse zum Zusammenhang von<br>g, Gf und Gc mit dem FFM aus vorherigen Studien .....               | 23         |
| 2.3.4 Betrachtungen zur Relevanz einer weiteren Metaanalyse zum<br>Zusammenhang zwischen Intelligenz und Persönlichkeit..... | 27         |
| 2.3 Publikationsbias.....  | 31         |
| <b>3 Zielsetzungen .....</b>   | <b>33</b>  |
| <b>4 Methode.....</b>  | <b>35</b>  |
| 4.1 Literaturrecherche.....  | 35         |
| 4.2 Studieninklusion und Zuordnung .....   | 35         |
| 4.3 Fixed Effects Model vs. Random Effects Model .....   | 39         |
| 4.4 Artefaktkorrekturen .....  | 41         |
| 4.5 Generalisierbarkeit von metaanalytischen Befunden .....  | 44         |
| 4.6 Spezifikation verwendeter Methoden zur Identifikation des<br>Publikationsbias .....                                      | 46         |
| 4.6.1 Identifikation und Adjustierung des Publikationsbias nach der<br>„trim and fill“-Methode .....                         | 46         |
| 4.6.2 Identifikation möglicher Publikationsverzerrungen mittels<br>Signifikanzfunktion im Streudiagramm .....                | 48         |
| 4.6.3 Index zum Verhältnis signifikanter und nichtsignifikanter<br>Effektgrößen .....  | 50         |

|   |            |
|---|------------|
| 4.6.4 Auswirkungen auf die Indexhöhe durch Parametervariationen.....  | 51         |
| 4.6.5 Bestimmung des Sensitivitätsbereichs zur Indexermittlung .....  | 58         |
| 4.6.7 Adaption der Simulation hinsichtlich der Stichprobengrößen und<br>-anzahl real publizierter Studien ..... | 60         |
| 4.6.8 Verwendete Methoden zur Untersuchung des Publikationsbias in der<br>vorliegenden Arbeit .....             | 64         |
| <b>5 Ergebnisse .....</b>   | <b>66</b>  |
| 5.1 Ergebnisse zur Allgemeinen Intelligenz und Neurotizismus .....  | 66         |
| 5.2 Ergebnisse zur Fluiden Intelligenz und Neurotizismus.....   | 76         |
| 5.3 Ergebnisse zur Kristallisierten Intelligenz und Neurotizismus .....   | 88         |
| 5.4 Ergebnisse zur Allgemeinen Intelligenz und Extraversion .....   | 95         |
| 5.5 Ergebnisse zur Fluiden Intelligenz und Extraversion.....  | 106        |
| 5.6 Ergebnisse zur Kristallisierten Intelligenz und Extraversion .....  | 115        |
| 5.7 Ergebnisse zur Allgemeinen Intelligenz und Offenheit für Erfahrungen.....                                   | 123        |
| 5.8 Ergebnisse zur Fluiden Intelligenz und Offenheit für Erfahrungen .....                                      | 134        |
| 5.9 Ergebnisse zur Kristallisierten Intelligenz und Offenheit für Erfahrungen.....                              | 144        |
| 5.10 Ergebnisse zur Allgemeinen Intelligenz und Verträglichkeit.....  | 154        |
| 5.11 Ergebnisse zur Fluiden Intelligenz und Verträglichkeit .....   | 164        |
| 5.12 Ergebnisse zur Kristallisierten Intelligenz und Verträglichkeit .....                                      | 170        |
| 5.13 Ergebnisse zur Allgemeinen Intelligenz und Gewissenhaftigkeit.....   | 179        |
| 5.14 Ergebnisse zur Fluiden Intelligenz und Gewissenhaftigkeit .....  | 190        |
| 5.15 Ergebnisse zur Kristallisierten Intelligenz und Gewissenhaftigkeit.....                                    | 200        |
| <b>6 Zusammenfassung und Diskussion.....</b>  | <b>209</b> |
| 6.1 Ergebnisse der theoretischen und methodischen Überlegungen .....  | 210        |
| 6.2 Ergebnisse und Diskussion zum Zusammenhang zwischen<br>Intelligenz und Persönlichkeit .....                 | 212        |
| 6.3 Grenzen der vorliegenden Arbeit und verbleibendes Arbeitsprogramm.....                                      | 230        |
| <b>Literatur.....</b>   | <b>234</b> |
| <b>Anhang .....</b>   | <b>260</b> |

## **Danksagung**

Mein größter Dank gilt Herrn Prof. Dr. André Beauducel, der mir die Möglichkeit eröffnet hat, eine Dissertationsschrift, trotz suboptimaler Umstände meinerseits, zu beginnen und als sein Mitarbeiter fertigzustellen. Fachlich und menschlich außergewöhnlich, adäquat fordernd und fördernd, dabei den Freiraum zur Bearbeitung der Thematik nicht restringierend, sei meinem Doktorvater herzlichst gedankt.

Ein ganz besonderer Dank geht an Frau Jun.-Prof. Dr. Anja Leue, die in zahlreichen Diskussionen keine Zeit und Mühe gescheut hat, mein Fachverständnis zu schärfen.

Nicht vermissen möchte ich die vielen intensiven, fruchtbaren, z. T. weit über das Fachliche hinausgehenden Diskussionen mit meinen ehemaligen Kolleginnen Dipl.-Psych. Anja Bath, Dipl.-Psych. Katharina Borch, Dipl.-Psych., Dipl.-Soz.-Päd. Elmira Marks, Dr. Madiha Rhana und Dr. Monika Thomsen.

Zu großem Dank fühle ich mich nicht zuletzt Herrn Dipl.-Psych. Frank Spohn verpflichtet, der durch seine gründliche, fachlich-kompetente und bescheidene Art in offenen, undogmatisch geführten Diskussionen wichtige Impulse für die vorliegende Arbeit setzte.

Bonn, im März 2013

## **1 Einleitung**

Auf der sechsten Tagung der European Association of Personality Psychology (1992) äußerte Kurt Pawlik: „Don't worry traits exist!“ (Westmeyer, S.748, 1995). Darauf erwiderte Hans Westmeyer, dass es Personen gebe, „die über einen privilegierten Zugang zum Eigentlichen dessen, was Persönlichkeit ausmacht, verfügen“ (Westmeyer, 1995, S. 748) und kommt zu dem Fazit: „Don't worry traitists exist!“ (Westmeyer, 1995, S. 752,). Diesem Diskurs liegt eine Debatte zwischen dem metaphysischen Realismus und dem sozialen Konstruktivismus zugrunde (vgl. Putnam, 1988). Diese Debatte ist zwar nicht Gegenstand der vorliegenden Arbeit, liefert jedoch einen Hintergrund, in dem die Ergebnisse der vorliegenden Arbeit einzuordnen sind. Vertreter des Fünf-Faktoren-Modells der Persönlichkeit, die scheinbar die tatsächliche Existenz der Persönlichkeitsdimensionen nicht in Frage stellen (bspw. McCrae & Costa, 1986; McCrae & John, 1992), könnten als metaphysische Realisten bezeichnet werden. Dabei kann es vorkommen, dass der rhetorischen Überzeugungskraft mehr Bedeutung beigemessen wird als dem empirischen Inhalt. Dazu ein Beispiel wie argumentiert werden kann, warum bzw. dass es fünf und nur fünf Persönlichkeitsdimensionen gibt: „We believe it is an empirical fact, like the fact that there are seven continents on earth or eight American presidents from Virginia“ (McCrae & John, 1992, S. 194). Der soziale Konstruktivismus unterscheidet sich vom metaphysischen Realismus insofern, dass keine tatsächliche Existenz im wissenschaftlichen Sinne angenommen wird, sondern dass sogenannte wissenschaftliche Tatsachen als von Menschen erfunden betrachtet werden (Scarr, 1985). Demnach kann angenommen werden, dass es keine Menge von Tatsachen gibt, die absolut und real sind. Für die vorliegende Arbeit, in der die fünf Persönlichkeitsdimensionen Neurotizismus, Extraversion, Offenheit für Erfahrungen, Verträglichkeit und Gewissenhaftigkeit im Zusammenhang mit Intelligenz metaanalytisch untersucht wurden, wurde weder eine tatsächliche Existenz von Persönlichkeitstraits angenommen, noch verworfen. Vielmehr wurde versucht, Veröffentlichungstendenzen und –trends hinsichtlich der Häufigkeit des Einsatzes des Fünf-Faktoren-Modells der Persönlichkeit, insbesondere der NEO-Inventare abzubilden.

Hinsichtlich der Fähigkeitskonstrukte der Allgemeinen, Fluiden und Kristallisierten Intelligenz wurde ebenfalls weder die Existenz von Traits angenommen noch verworfen. Von der Einschätzung ausgehend, dass die Autoren der publizierten Studien eine z.T. liberalere bzw. ungenauere Zuordnung der jeweiligen Intelligenztests bzw. kognitiven Leistungsaufgaben zu den Fähigkeitskonstrukten vorgenommen haben, wurde die Klassifizierung der Fähigkeitskonstrukte, wie sie von den Autoren der publizierten Studien erfolgte, beibehalten. Damit wurde versucht, ein sozialkonstruktivistisches Wissenschaftsverständnis in der Metaanalyse abzubilden. Darüber



## *Kapitel 1 – Einleitung*

hinaus schien es auch angemessen, Befunde auf Basis einer einheitlichen Zuordnung zu generieren. Dazu wurde in einer zweiten Analyse-Variante, eine nachträgliche Zuordnung der Fähigkeitskonstrukte gemäß einer einheitlichen Intelligenzstrukturtheorie durch den Autor der vorliegenden Arbeit durchgeführt. Ein Vergleich der beiden Analyse-Varianten erlaubt eine Einschätzung der Auswirkungen unterschiedlicher Zuordnungen der Fähigkeitskonstrukte auf die metaanalytischen Kennwerte.

Die als Publikationsbias bezeichnete Tendenz, bevorzugt signifikante Effektgrößen zu publizieren, führt zu einer Überschätzung des Populationszusammenhangs und letztlich zu einer eingeschränkten Glaubwürdigkeit der Forschung (Ferguson & Heene, 2012). Dabei scheinen die Untersuchungen zur Identifikation eines Publikationsbias in Metaanalysen von großem Interesse und Relevanz zu sein (bspw. Coursol & Wagner, 1986; Duval & Tweedie, 2000a, 2000b; Ferguson & Brannick, 2012; Ferguson & Heene, 2012; Hedges & Olkin, 1985; Hunter & Schmidt, 2004; Light & Pillemer, 1984; Rosenthal, 1979; Rothstein & Bushman, 2012). Ein Publikationsbias wurde im Kontext des Zusammenhangs zwischen Intelligenz und Persönlichkeit bisher noch nicht untersucht. Dennoch kann es als wichtig erachtet werden zu wissen, inwieweit die Höhe des Populationszusammenhangs zwischen Intelligenz und Persönlichkeit auf einer bevorzugten Publikation signifikanter Effektgrößen beruht. Deshalb wurde im Rahmen der vorliegenden Arbeit der Publikationsbias untersucht.

**Zum Aufbau der Arbeit**

In Kapitel 2 erfolgt die theoretische Einordnung und Relevanz der Fähigkeitskonstrukte der Allgemeinen, Fluiden und Kristallisierten Intelligenz sowie der Persönlichkeitsdimensionen Neurotizismus, Extraversion, Offenheit für Erfahrung, Verträglichkeit und Gewissenhaftigkeit. Anschließend werden unterschiedliche Perspektiven bzw. Modelle sowie bisherige metaanalytische Ergebnisse zum Zusammenhang zwischen Intelligenz und Persönlichkeit dargestellt. Darüber hinaus werden verschiedene Aspekte des Publikationsbias erläutert. Aus der theoretischen Einordnung und Relevanz einer Metaanalyse und Untersuchung eines Publikationsbias werden in Kapitel 3 die Ziele der vorliegenden Arbeit artikuliert. Die methodischen Überlegungen zur Durchführung einer Metaanalyse sowie der Einsatz verschiedener Methoden zur Identifikation des Publikationsbias sind in Kapitel 4 beschrieben. In Kapitel 5 sind die metaanalytischen Ergebnisse zum Zusammenhang zwischen den oben genannten Fähigkeitskonstrukten und Persönlichkeitsdimensionen, einschließlich der Resultate zum Publikationsbias, dargestellt. In Kapitel 6 erfolgt schließlich eine zusammenfassende Diskussion, welche mit den Grenzen der vorliegenden Arbeit und dem weiteren Forschungsbedarf endet.

## 2 Theorie

### 2.1 Für die vorliegende Arbeit relevante Intelligenzstrukturmodelle

Das hypothetische Konstrukt Intelligenz ist durch eine Vielzahl von kognitiven Teilfähigkeiten gekennzeichnet. Im Zuge der Forschungsentwicklung werden diese Teilfähigkeiten immer weiter erschlossen. Man bezeichnet hypothetische Konstrukte wegen ihrer prinzipiellen Erweiterungsfähigkeit und permanenten Erweiterung auch als offene Konstrukte (Brocke & Beauducel, 2001; Amthauer, Brocke, Liepmann & Beauducel, 2001). Deshalb lässt sich eine enge, allgemeingültige Definition des Begriffes Intelligenz nur schwer formulieren. Eine allgemeingültige Definition könnte sein: Intelligenz umfasst ausschließlich kognitive Fähigkeiten. Dabei wird eine Ausweitung auf andere Fähigkeiten abgelehnt (Schuler, 2002).

Gemeinsamkeiten vieler kognitiver Leistungstests (kogLT) sind aufsteigende Aufgabenschwierigkeiten, Zeitbegrenzungen, oft die Vorgabe verbaler, numerischer und räumlicher Aufgaben. Dabei hält auf der einen Seite eine Gruppe von Forschern daran fest, dass die Intelligenz auf einen allgemeinen Intelligenzfaktor „g“ reduziert werden kann (bspw. Spearman 1904; Vernon, 1987; Jensen, 1998). Auf der anderen Seite gibt es integrative Strukturmodelle, die spezifische und generelle Fähigkeitskonstrukte zu integrieren versuchen (Süß & Beauducel, 2011). Hierzu zählen das Berliner Intelligenzstrukturmodell von Jäger (1982), die Three-Stratum Theory (TST) von Carroll (1993, 2005), die Cattell-Horn-Carroll-Theorie (CHC-Theorie) von McGrew (1997, 2005, 2009) und mit Einschränkungen die erweiterte Theorie der fluiden und kristallisierten Intelligenz (erweiterte Gf-Gc-Theorie) von (Horn & Blankson, 2005; Horn & Noll, 1997), in der allerdings die Annahme eines g-Faktors abgelehnt wird.

Die Theorie der fluiden und kristallisierten Intelligenz (Gf-Gc Theorie; Cattell, 1963; Cattell, 1971; Horn & Blankson, 2005; Horn & Cattell, 1966, 1967; Horn & Noll, 1997; Horn & Stankov, 1982) ist vermutlich die einflussreichste Strukturtheorie der Intelligenz (Ackerman, 1996; Baltes, 1987, 1997; Stankov, 2000; Süß & Beauducel, 2011). Dabei soll Gf einen „process of reasoning and other mental activities that depend only minimally on learning and acculturation“ abbilden (Carroll, 1993, S. 624). Hingegen werden mit Gc Fähigkeiten bezeichnet, die primär auf den Einfluss der Akkulturation bzw. auf Bildung zurückgeführt werden können (Cattell, 1987; Horn, 1988).

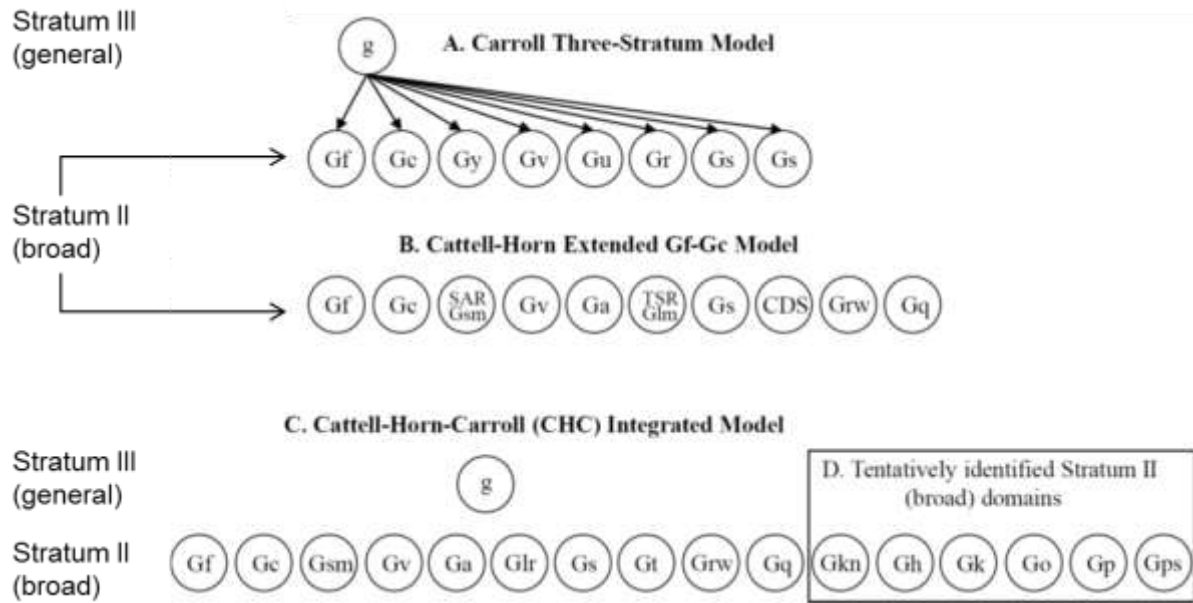
In der erweiterten Gf-Gc-Theorie werden 10 Faktoren zweiter Ordnung (Sekundärfaktoren) angenommen (Horn & Blankson, 2005; Horn & Noll, 1997). Neben Gf und Gc sind folgende

Fähigkeitskonstrukte eingegangen: Kurzeitgedächtnis (*short-term apprehension and retrieval*), Verarbeitung visueller Informationen (*visual processing*), Verarbeitung auditiver Informationen (*auditive processing*), Flüssigkeit des Abrufs von Informationen aus dem Langzeitgedächtnis (*fluency of retrieval from long-term storage*), Information, Verarbeitungsgeschwindigkeit (*processing speed*) Geschwindigkeit für richtige Entscheidungen (*correct decision speed*), Lesen und Schreiben (*reading and writing*), quantitatives Wissen (*quantitative knowledge*; siehe auch Abbildung 1).

Eine der umfangreichsten Entwürfe zur Integrierbarkeit von unterschiedlichen Intelligenzstrukturmodellen stellt die Arbeit von Carroll (1993) dar (Horn & Noll, 1997; McGrew, 2009; Süß, 2001; Süß & Beauducel, 2011). Der „Three-Stratum Theory“ (TST; Carroll, 1993) liegen 461 Datensätze zugrunde, die mittels einer einheitlichen faktorenanalytischen Methodik reanalysiert und die Befunde ähnlich einem Puzzle zu einem hierarchischen Strukturmodell zusammengesetzt wurden (Süß & Beauducel, 2011). Dabei wird von Carroll (1993) auf dem obersten, dritten Stratum ein Faktor *g* angenommen. Im zweiten Stratum sind folgende Fähigkeitskonstrukte postuliert: *Gf*, *Gc*, allgemeines Gedächtnis und Lernen (*general memory and learning*), visuelle Wahrnehmung (*visual processing*), auditive Wahrnehmung (*broad visual perception*), Wiedergabefähigkeit (*broad retrieval ability*), kognitive Schnelligkeit (*broad cognitive speediness*), Verarbeitungsgeschwindigkeit im Sinne von Reaktionszeiten (*processing speed*; siehe Abbildung 1). Dabei sollen *Gf* und *Gc* die höchste *g*-Sättigung auf dem zweiten Stratum haben (Carroll, 1993). Im ersten Stratum werden 68 spezifische Fähigkeitskonstrukte angenommen (Carroll, 1993).

Ein wesentlicher Unterschied zwischen der erweiterten *Gf-Gc*-Theorie und der TST ist, dass in der TST auf dem dritten Stratum ein genereller Intelligenzfaktor (*g*) angenommen wird. Berichtet Carroll (1993) „abundant evidence“ für ein Konstrukt von *g* aufgrund seiner Befunde, lehnt Horn (2008) ein generelles Konstrukt *g* zur Erklärung der Interkorrelationen der Faktoren zweiter Ordnung ab. Gegen die Annahme eines *g*-Faktors wird von Horn (2008) sowie Horn und Blankson (2005) angeführt, dass die einzelnen Konstrukte durch die jeweils neun anderen nicht vorhergesagt werden können. Dies widerspreche der Annahme gleichartiger zugrunde liegender Prozesse. Überdies werden von Horn (2008) die inzwischen gut gesicherten substantiellen Unterschiede im Entwicklungsverlauf der 10 Fähigkeiten über die Lebensspanne angeführt (Horn, 1994, 1998; Lindenberger & Baltes, 1997; Mayer & Baltes, 1996; Salthouse, 2009; Schaie, 1994).

Ein weiterer Unterschied ist die Inklusion des Fähigkeitskonstrukts quantitatives Wissen auf dem Stratum II in der erweiterten Gf-Gc-Theorie, jedoch nicht in der TST (Abbildung 1). In der TST wird das Fähigkeitskonstrukt Lesen und Schreiben unter Gc subsumiert und in der erweiterten Gf-Gc-Theorie separat im zweiten Stratum aufgeführt. Das Fähigkeitskonstrukt allgemeines Gedächtnis und Lernen auf dem zweiten Stratum in der TST wird in der Gf-Gc-Theorie unterteilt in Kurzzeitgedächtnis und Langzeitgedächtnis (McGrew, 2009; Abbildung 1).



**CHC Broad (Stratum II) Ability Domains**

|     |                                 |     |                                     |
|-----|---------------------------------|-----|-------------------------------------|
| Gf  | Fluid reasoning                 | Grw | Reading and writing                 |
| Gc  | Comprehension-knowledge         | Gq  | Quantitative knowledge              |
| Gsm | Short-term memory               | Gkn | General (domain-specific) knowledge |
| Gv  | Visual processing               | Gh  | Tactile abilities                   |
| Ga  | Auditory processing             | Gk  | Kinesthetic abilities               |
| Glr | Long-term storage and retrieval | Go  | Olfactory abilities                 |
| Gs  | Cognitive processing speed      | Gp  | Psychomotor abilities               |
| Gt  | Decision and reading speed      | Gps | Psychomotor speed                   |

**Abbildung 1.** Schematische Darstellung und Vergleich der Three-Stratum Theorie von Carroll, der erweiterten Gf-Gc Theorie von Cattell-Horn und der integrierten Cattell-Horn-Carroll Theorie nach McGrew (2009, S. 4); Das erste Stratum ist aufgrund der Übersichtlichkeit hier nicht abgebildet.

Die Cattell-Horn-Carroll-Theorie (CHC-Theorie; McGrew, 2005, 2009) kann als eine Synthese zwischen der TST (Carroll, 1993) und der erweiterten Gf-Gc-Theorie (Horn & Blankson, 2005; Horn & Noll, 1997) aufgefasst werden (McGrew, 2009; Süß & Beauducel, 2011). Die CHC-Theorie nimmt ein hierarchisches Drei-Ebenenmodell an. Auf der obersten Ebene wird ein allgemeiner Intelligenzfaktor angenommen, auf der zweiten Ebene werden 10 Fähigkeitskonstrukte aus der erweiterten Gf-Gc-Theorie postuliert. Überdies werden auf der zweiten Ebene sechs weitere Faktoren als tentative, noch nicht gesicherte Faktoren angenommen:

Allgemeines, bereichsspezifisches) Wissen (*general, domain-specific knowledge*), taktile Fähigkeiten (*tactile abilities*), kinesthetische Fähigkeiten (*kinesthetic abilities*), olfaktorische Fähigkeiten (*olfactory abilities*), psychomotorische Fähigkeiten (*psychomotor abilities*), psychomotorische Geschwindigkeit (*psychomotor speed*; siehe Abbildung 1). Diese sechs Faktoren wurden bereits von Carroll (1993) als vorläufige und noch ungesicherte Fähigkeitskonstrukte beschrieben. Darüber hinaus sollen faktorenanalytische Ergebnisse diese sechs Fähigkeitskonstrukte auf der zweiten Ebene bestätigen (McGrew, 2005; 2009). Auf der dritten Ebene werden nach der CHC-Theorie die Primärfaktoren der TST angenommen (Süß & Beauducel, 2011). Für ausführliche Darstellungen zur historischen Entwicklung der Intelligenzmodelle sowie zum Zusammenhang zwischen Intelligenzstrukturmodellen und Intelligenzmessungen sei auf weiterführende Literatur verwiesen (bspw. Carroll, 1982, 1993; Süß & Beauducel, 2011).

Da sowohl in der TST als auch in der CHC-Theorie die Annahme eines g-Faktors postuliert wird, und eine Vielzahl von Studien, insbesondere auch in Metaanalysen (Ackerman & Heggestad, 1997; Judge, Jackson, Shaw, Scott & Rich, 2007; Poropat, 2009; Wolf & Ackerman, 2005) Zusammenhänge mit g untersucht haben, ist der g-Faktor relevant für die vorliegende metaanalytische Arbeit. Da die beiden Fähigkeitskonstrukte Gf und Gc in der TST, der erweiterten Gf-Gc-Theorie und der CHC-Theorie Berücksichtigung finden und somit grundsätzlich als relevante Faktoren im Kontext hierarchischer Modells der Intelligenz betrachtet werden (Beauducel, Brocke & Liepmann, 2001, Carroll, 1993; Härnqvist, Gustafsson, Muthén & Nelson, 1994; Horn & Blankson, 2005; Horn & Noll, 1997; McGrew, 1997, 2005, 2009; Süß & Beauducel, 2011), sollen Gf und Gc ebenfalls in der vorliegenden Arbeit berücksichtigt werden.

Neben der theoretischen und empirischen Relevanz von g, Gf und Gc in der Intelligenzstrukturforschung scheinen die genannten Intelligenzkonstrukte auch praktische Relevanz zu haben. So berichtet Süß (2001), dass die Verarbeitungskapazität aus dem Berliner Intelligenzstrukturmodell von Jäger (1982), welche eine hohe Ähnlichkeit mit Gf aufweist, einen der besten Einzelprädiktoren für die Prognosen interindividueller Differenzen bei kognitiven Leistungen und Kriterien im alltäglichen Leben darstellt.

In einer Arbeit, die unterschiedliche metaanalytische Befunde zur Validität eignungsdiagnostischer Verfahren im beruflichen Kontext untersucht (Schmidt & Hunter, 1998), wird ein artefaktkorrigierter Validitätskoeffizient von .51 bei der Vorhersage von Berufserfolg durch g

berichtet, der lediglich durch einen Validitätskoeffizienten von .54 durch Arbeitsproben übertroffen wird. Mit Validitätskoeffizienten von .48 (Schmidt & Hunter, 1998) und .45 (Dye, Reck & McDaniel, 1993) für einen Fachkenntnistest, bei dem angenommen werden kann, dass dieser hoch auf Gc lädt, liegt dieses diagnostische Verfahren ebenfalls an der Spitze, um Berufserfolg vorherzusagen.

Für die Intelligenzkonstrukte g, Gf und Gc kann angenommen werden, dass sie einen relevanten Ausschnitt aus der Intelligenzstrukturforschung abbilden (Carroll, 1993; Horn & Blankson, 2005; Horn & Noll, 1997; McGrew, 1997, 2005, 2009; Süß & Beauducel, 2011). Gleichzeitig kann die praktische Relevanz von g, Gf und Gc, gestützt durch eine Vielzahl gut gesicherter substantieller Befunde, angenommen werden (Dye et al., 1993; Schmidt & Hunter, 1998; Judge et al., 2007; Liepmann, Beauducel, Brocke & Amthauer, 2008). Diese vermutlich einflussreichsten Fähigkeitskonstrukte in der Intelligenzstrukturforschung zeigen eine zunehmende Konvergenz über verschiedene Modelle und Anwendungen hinweg (siehe oben) und sind deshalb für eine metaanalytische Untersuchung besonders geeignet. Durch eine zunehmende Konvergenz der Fähigkeitskonstrukte g, Gf und Gc, verringert sich das Inkommensurabilitätsproblem (Glass & Kliegl, 1983). „Äpfel“ können so unter dem Konstrukt „Apfel“ und „Birnen“ unter dem Konstrukt „Birne“ aggregiert werden und müssen in dem Fall nicht ausschließlich unter dem Konstrukt „Kernobst“ subsumiert werden (vgl. Wittmann, 1985). Das Aggregationsniveau zur Bestimmung metaanalytischer Kennwerte unterliegt weniger Subsumierungen, was möglicherweise die Robustheit der Ergebnisse erhöht.

## **2.2 Fünf-Faktoren-Modell der Persönlichkeit**

Die Forschung zum Fünf-Faktoren-Modell der Persönlichkeit (FFM) hat ihre Anfänge in den 20er Jahren des 20. Jahrhunderts. In der Sedimentationshypothese (Klages, 1926), auch als lexikalischer Ansatz bezeichnet, wird angenommen, dass sich alle interindividuellen Unterschiede in der natürlichen Sprache wiederfinden. Diese Annahme berücksichtigend, analysierten Allport und Odbert (1936) Wörter aus dem „Webster’s Dictionary“ mit denen sich Persönlichkeit beschreiben lässt. Die 4,500 persönlichkeitsbeschreibenden Begriffe aus der Liste von Allport und Odbert (1936) wurde von Cattell (1943, 1945) auf 12 persönlichkeitspezifische Faktoren reduziert. Eine weitere Reduzierung auf fünf Faktoren erfolgte erstmals durch Fiske (1949) sowie Tupes und Christal (1958, 1961) und bildet damit den eigentlichen Ausgangspunkt des heutigen FFM im engeren Sinne. Zur Bestimmung von Persönlichkeitsvariablen untersuchten Tupes und Christal (1958, 1961) Korrelationsmatrizen aus acht verschiedenen Stichproben in denen sich konsistent

fünf Faktoren gefunden haben sollen. Weitere Untersuchungen folgten u.a. von Norman (1963), der ebenfalls berichtete, fünf Persönlichkeitsfaktoren gefunden zu haben.

Für die fünf Persönlichkeitsfaktoren wurden verschiedene Bezeichnungen mit z.T. unterschiedlichen Akzentuierungen gewählt. So bezeichnet bspw. Fiske (1949) den Faktor V als „Inquiring Intellect“. Hingegen wurde dieser Faktor von Norman (1963) als „Culture“ bezeichnet. In nachfolgenden Untersuchungen wurde der Faktor V als „Intellectual Interest“ (Goldberg, 1981, 1990) oder auch als „Openness to Experience“ (Costa & McCrae, 1985) bezeichnet. Aufgrund der Vielzahl verschiedener Untersuchungen und Entwicklungen (bspw. Costa & McCrae, 1985; Goldberg, 1982; Norman, 1963; Peabody, 1987; Tupes & Christal, 1958, 1961) scheint es unzureichend zu sein, einen einzelnen Forscher bzw. eine einzelne Forschergruppe als Begründer des FFM zu bezeichnen. Für ausführliche Darstellungen der Geschichte und Entwicklung des FFM sei bspw. auf Digman (1990, 1996) oder John und Srivastava (1999) verwiesen.

Erst in den 1980er Jahren wurden die ersten Inventare zur Erfassung des FFM veröffentlicht. Im Jahr 1985 wurde das „NEO-Personality-Inventory“ (NEO-PI) im englischsprachigen Raum zur Erfassung der Persönlichkeitsdimensionen „Neuroticism“, „Extraversion“, „Openness to Experience“, „Agreeableness“ und „Conscientiousness“ publiziert. 1989 folgte das NEO-Five-Factor-Inventory (NEO-FFI) und 1992 das NEO-PI in einer revidierten Fassung (NEO-PI-R) von Costa und McCrae. Offenbar hat die Vermarktung der NEO-Inventare von Costa und McCrae zur Popularität des FFM beigetragen. Darüber hinaus kann mit Beginn der 1990er Jahre eine Dominanz des FFM und der NEO-Inventare in der Persönlichkeitsforschung angenommen werden (Schulze & Roberts, 2006). Diese Popularität des FFM scheint einerseits auf die Hoffnung und Überzeugung zurückzugehen, mit Hilfe faktorenanalytisch begründeter fünf Persönlichkeitsfaktoren höherer Ordnung ein Gesamtsystem der Persönlichkeit ermittelt zu haben (Amelang, Bartussek, Stemmler & Hagemann, 2006). Andererseits scheint der ökonomische Einsatz, insbesondere der NEO-Verfahren, die leichte Verständlichkeit der fünf Dimensionen sowie die guten bis mäßigen Güteeigenschaften zur Popularität des FFM beizutragen (Ostendorf & Angleitner, 2004).

Zur Popularität des FFM haben auch die für die Praxis relevanten Untersuchungen beigetragen. So wurden u.a. im beruflichen Kontext die Persönlichkeitsdimensionen des FFM in einer Vielzahl von Studien als Prädiktor zur Vorhersage von Arbeitsleistung oder Berufszufriedenheit berücksichtigt (bspw. Barrick, Mount & Judge, 2001; Barrick, Steward & Piotrowski, 2002;



Blickle, Wendel & Ferris, 2010; De Fruyt & Mervielde, 1996; Hurtz & Donovan, 2000; Judge, Heller & Mount, 2002; Judge & Ilies, 2002). Metaanalytischen Ergebnissen zu Folge kann angenommen werden, dass sich die Arbeitsleistung durch die Persönlichkeitsdimension Gewissenhaftigkeit mit einem artefaktkorrigierten Validitätskoeffizienten von .24 vorhersagen lässt (Barrick et al., 2001) oder dass die Berufszufriedenheit durch Neurotizismus mit einem artefaktkorrigierten Validitätskoeffizienten von -.24 vorhergesagt werden kann (Judge et al., 2002).

Der Anspruch, dass das FFM ein überlegenes Persönlichkeitsmodell sein soll, kommt im Invarianz-Postulat und Comprehensiveness-Postulat zum Ausdruck (bspw. McCrae & John, 1992; McCrae & Costa, 1986). Beim Invarianz-Postulat wird angenommen, dass eine Robustheit des FFM gegeben ist, unabhängig von:

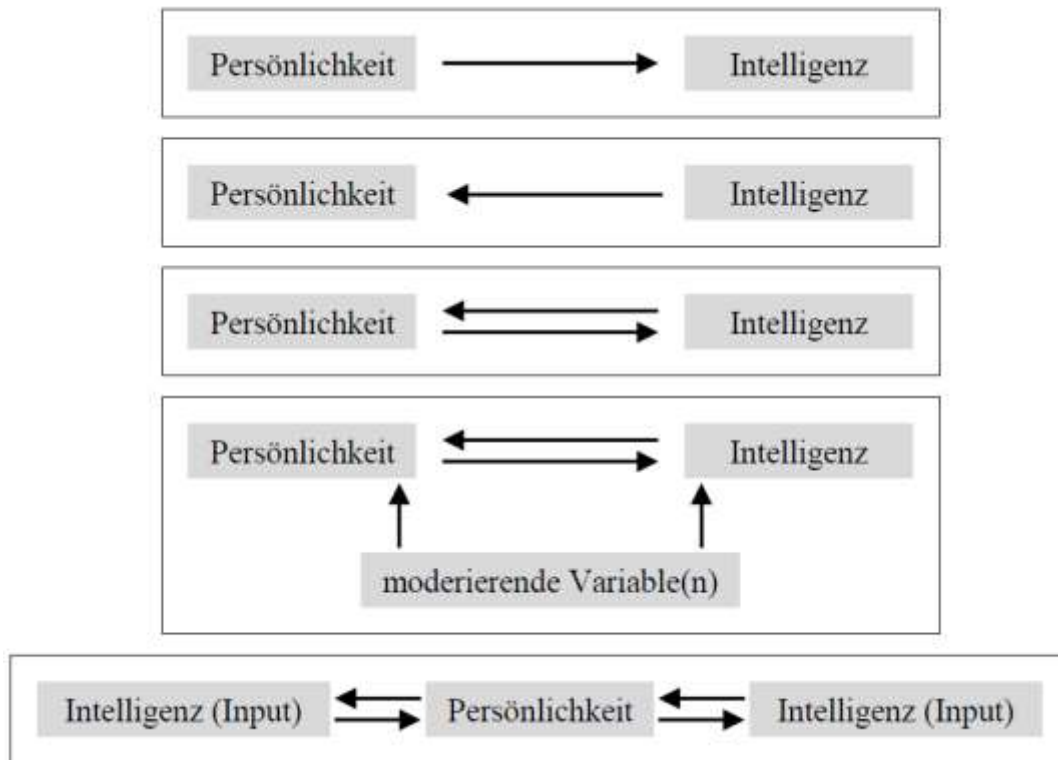
- Variationen zwischen (interkulturellen) Probandengruppen,
- faktorenanalytischen Methoden,
- semantischen Zugängen (bspw. Sätze vs. Adjektive) und
- verschiedenen diagnostisch-instrumentellen Ansätzen (bspw. Fremd- vs. Selbsteinschätzung).

Beim Comprehensiveness-Postulat wird ein Vollständigkeitsanspruch bzw. eine Erklärungspotenz angenommen. Dabei soll sich die Persönlichkeit mit den Persönlichkeitsdimensionen erschöpfend beschreiben lassen (McCrae & John, 1992; McCrae & Costa, 1986). Dennoch gibt es sowohl zum Invarianz-Postulat als auch Comprehensiveness-Postulat kritische Anmerkungen (bspw. Andresen, 2001; Brocke, 2000; Block, 1995; Brody & Ehrlichman, 1998; Peabody, 1987; Pervin, 1994).

Wie bereits in der Einleitung beschrieben, soll das FFM hinsichtlich des Geltungsanspruchs in der vorliegenden Arbeit weder kritisiert noch gelobt werden. Durch die Sichtung publizierter Studien zum Zusammenhang zwischen Intelligenz und Persönlichkeit aus den Jahren von 1985 bis 2011 zeigte sich, dass in 67.86% der Studien NEO-Inventare zur Erfassung der Persönlichkeit verwendet wurden. Das lässt auf die Popularität des FFM und insbesondere auf die Popularität von NEO-Inventaren schließen. Wie im Intelligenzbereich tritt auch im Persönlichkeitsbereich eine zunehmende Konvergenz der Konstrukte auf. Dies macht eine metaanalytische Untersuchung zum Zusammenhang zwischen Intelligenz und Persönlichkeit besonders interessant, da sich komplexe Integrationen der Intelligenz- und Persönlichkeitskonzepte, wie bspw. in Ackerman und Heggestad (1997), verringern. Durch weniger Subsumierungen kann davon ausgegangen werden, dass die Robustheit der metaanalytischen Kennwerte zunimmt.

### 2.3 Zusammenhang von Intelligenz und Persönlichkeit

Im Kontext des Zusammenhangs zwischen Intelligenz und Persönlichkeit können anhand theoriegeleiteter Annahmen, empirischer Befunde sowie konzeptioneller Eigenheiten unterschiedliche Perspektiven eingenommen werden. Mögliche Richtungen des Zusammenwirkens von Intelligenz und Persönlichkeit sind schematisch in Abbildung 2 nach Zeidner (1995) dargestellt.



**Abbildung 2.** Visualisierung theoretischer Annahmen, in welche Richtung(en) sich Intelligenz und Persönlichkeit beeinflussen können nach Zeidner (1995, S. 316).

Neben den fünf Perspektiven des Zusammenwirkens von Intelligenz und Persönlichkeit unterscheiden Stumm, Chamorro-Premuzic & Ackerman (2011) drei unterschiedliche Perspektiven (Abbildung 3). Dabei wird in der ersten Perspektive eine konzeptionelle und empirisch weitestgehende Unabhängigkeit von Intelligenz und Persönlichkeit angenommen. In einer zweiten Perspektive wird eine Beeinflussung der Intelligenzleistung durch die Persönlichkeit angenommen, in der Persönlichkeitsdimensionen, wie z.B. Neurotizismus, die Leistungen in Intelligenztests determinieren (vgl. „Top-Down“-Ansatz; Chamorro-Premuzic & Furnham, 2004, Rindermann & Neubauer, 2001). In einer dritten Perspektive wird berichtet, dass die Persönlichkeit einer Person darüber entscheidet wie, wann, wo und wieviel in kognitive Aktivitäten investiert wird. Folglich wird eine Determination der Intelligenz durch die Persönlichkeit über die Zeit angenommen (Stumm et al., 2011). Wurde in den 1990er Jahren eine

Determination der Persönlichkeit durch kognitive Fähigkeiten diskutiert (Zeidner, 1995), scheint sich im Laufe der Zeit diese Perspektive nicht etabliert zu haben (Stumm et al., 2011).



**Abbildung 3.** Theoretische Perspektiven möglicher Intelligenz- und Persönlichkeits Assoziationen nach Stumm et al. (2011, S. 219).

Um die verschiedenen Perspektiven des Zusammenhangs zwischen Intelligenz und Persönlichkeit besser einordnen zu können, sollen im Folgenden

- Gemeinsamkeiten und Unterschiede der beiden Konstrukte beschrieben werden (Abschnitt 2.3.1),
- prominente Modelle erläutert werden, die das Zusammenwirken von Intelligenz und Persönlichkeit erklären könnten (Abschnitt 2.3.2) und
- in der Literatur publizierte metaanalytische Kennwerte zum Zusammenhang zwischen Intelligenz und Persönlichkeit berichtet werden (Abschnitt 2.3.3).

### 2.3.1 Gemeinsamkeiten und Unterschiede von Intelligenz und Persönlichkeit

Die Konstrukte der Intelligenz und der Persönlichkeit weisen eine Vielzahl von Gemeinsamkeiten, aber auch wesentliche Unterschiede auf. Gemeinsamkeiten sind:

- Sowohl bei der Intelligenz als auch bei der Persönlichkeit wird eine zeitliche Stabilität angenommen (bspw. Caspi, 2000; Deary, Whalley, Lemmon, Crawford & Starr, 2000).

- Beide Konstrukte sind quantifizierbar durch psychometrische Tests und sollen genetisch, wenn auch unterschiedlich, determiniert sein (bspw. Funder, 2001; Spinath & Johnson, 2011).
- KogLt und Persönlichkeitstests werden im Kontext der Untersuchung individueller Unterschiede eingesetzt (Stumm et al., 2011).
- Mit kogLt und Persönlichkeitstests können empirische Prognosen in unterschiedlichen Kontexten vorhergesagt werden (bspw. Barrick et al., 2001; Judge et al., 2002; Poropat, 2009; Salgado et al., 2003; Schmidt & Hunter, 1998).

Neben den Gemeinsamkeiten von Intelligenz und Persönlichkeit gibt es auch Unterschiede. Bei der Erfassung der Intelligenz durch kogLt soll die maximale kognitive Performanz erfasst werden. Hingegen sollen bei der Persönlichkeit mit Selbst- oder Fremdeinschätzungen generelle Verhaltensweisen, wie sich eine Person typischerweise in einer bestimmten Situation verhält, erfasst werden. Die Unterschiede, die sich durch die Operationalisierung der maximalen kognitiven Performanz vs. Selbst- oder Fremdeinschätzung genereller Verhaltensweisen ergeben, sollen nachfolgend kurz beschrieben werden.

Eine Person kann sich bei der Bearbeitung eines kogLt bzw. bei der Erfassung der maximalen kognitiven Performanz nicht „intelligenter machen“ als sie tatsächlich ist. Allerdings kann eine Person in kogLt absichtlich schlechtere Testresultate erzielen. Dabei soll die Tendenz, absichtlich schlechtere Testresultate zu erzielen, gering ausgeprägt sein, da es gesellschaftlich nicht als erstrebenswert gilt, einen niedrigen IQ-Wert zurückgemeldet zu bekommen (Zeidner & Matthews, 2000). Darüber hinaus werden kogLt häufig in prüfungs- bzw. karriererelevanten Kontexten eingesetzt. Eine absichtliche Minderung der kognitiven Performanz würde zu Benachteiligungen gegenüber Mitbewerbern führen und scheint daher ebenfalls als eher unwahrscheinlich (Zeidner & Matthews, 2000).

Mit dem Einsatz von Persönlichkeitsinventaren sollen durch Selbst- bzw. Fremdeinschätzungen generelle Verhaltensweisen erfasst werden, um zu persönlichkeitsrelevanten Erkenntnissen zu gelangen. Während der Bearbeitung von Persönlichkeitstests bedarf es keiner extremen Ausprägung, um einem normativ erwünschten Maximum zu entsprechen. Persönlichkeitsinventare sind daher bidirektional ausgelegt (Tabelle 1) und sollen das typische Verhalten in Alltagssituationen erfassen (Most & Zeidner, 1995). Ein selbst- oder fremdberichtetes Persönlichkeitsinventar hat nach Zeidner und Matthews (2000) kein objektives Kriterium für die

gegebenen Antworten im Vergleich zum kogLt. Dadurch besteht eine größere Gefahr der Verzerrung durch Antworttendenzen (bspw. soziale Erwünschtheit, Akquieszenz, Tendenz zur Mitte, Tendenz zu extremen Antwortkategorien). Selbst wenn die Absicht besteht, sich wahrheitsgemäß zu äußern, sind die selbst- oder fremdberichteten Persönlichkeitsinventare von dem Eindruck bzw. der Wahrnehmung der Person abhängig, die das Persönlichkeitsinventar ausfüllt. Dementsprechend ist es schwer zu unterscheiden, ob eine Person keine korrekten Antworten geben möchte oder aufgrund einer unzureichenden Fähigkeit zur Selbstreflexion dazu nicht in der Lage ist (Thorndike, 1982). Darüber hinaus kann nicht ausgeschlossen werden, dass das Ausfüllen eines selbst- bzw. fremdberichteten Persönlichkeitsinventars stärker durch Stimmungen und emotionale Zustände beeinflusst ist als die Bearbeitung eines kognitiven Leistungstests (Zeidner & Matthews, 2000).

**Tabelle 1**

Unterschiedliche Aspekte und ihre Konzeptsdifferenzierungen bei der Operationalisierung der Intelligenz mit kognitiven Leistungstests und der Persönlichkeit mit selbstberichteten Persönlichkeitsinventaren, angepasst nach Most und Zeidner (1995, S. 583)

| <b>Dimension</b>                         | <b>Intelligenz</b>                        | <b>Persönlichkeit</b>   |
|--|---|---|
| Trait                                    | Unidirektional<br>(wenig bis viel)        | Bidirektional<br>(entgegengesetzte Extreme)                     |
| Situation und Ziele                      | abstrakte Situation,<br>Maximalperformanz | alltägliche Situationen,<br>typische Verhaltensweisen           |
| Instruktion                              | Versuch der besten Leistung               | ehrliche Antwort  |
| Kriterium zur Erfassung<br>der Antworten | objektives Kriterium                      | Richtung und Intensität,<br>(es gibt keine richtige<br>Antwort) |
| Reliabilität des Inventars               | hoch                                      | variiert von hoch bis niedrig                                   |
| Interpretation der Resultate             | relativ direkt                            | offener und kontroverser  |

Ein kognitiver Leistungstest unterliegt einer deutlich geringeren Kontrolle als ein Persönlichkeitsfragebogen, da es in der Regel nur eine richtige Antwort gibt, und die standardisierten Aufgaben in einer begrenzten Zeit bearbeitet werden müssen. Dementsprechend ist die Möglichkeit der Verfälschbarkeit von kogLt als geringer zu bewerten im Vergleich zu selbst- oder fremdberichteten Persönlichkeitsinventaren. Für eine ausführliche Darstellung von Gemeinsamkeiten und Unterschieden zur Operationalisierung der kognitiven Leistungsfähigkeit durch kogLt und zur Operationalisierung der Persönlichkeit durch selbst- oder fremdberichtete Persönlichkeitsinventare sei bspw. auf die weiterführende Literatur von Chamorro-Premuzic,

Stumm und Furnham (2011), Most und Zeidner (1995) sowie Zeidner und Matthews (2000) verwiesen.

### **2.3.2 Modelle zum Zusammenwirken von Intelligenz und Persönlichkeit**

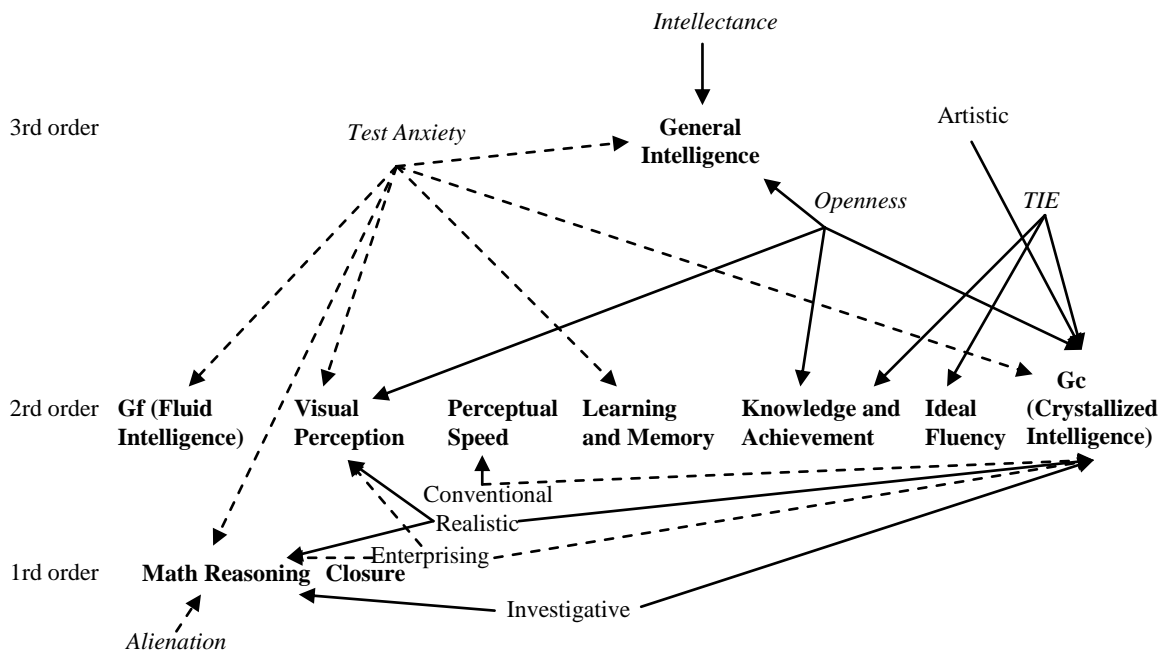
Das Bemühen, menschliches Verhalten mit Hilfe von Intelligenz und Persönlichkeit erklären zu wollen, weist eine lange Forschungsgeschichte auf (Saklofske & Zeidner, 1995). Dabei scheint das Interesse, Modelle über das Zusammenwirken von Intelligenz und Persönlichkeit abzubilden, bis in die Gegenwart vorhanden zu sein (bspw. Beauducel, Liepmann, Felfe & Nettelstroth, 2007; Chamorro-Premuzic & Arteche, 2008, Stumm et al., 2011).

Modelle bieten den Vorteil, einen komplexen Sachverhalt zu vereinfachen bzw. den Inhalt auf das Wesentliche zu beschränken. Dennoch weisen Stankov, Boyle und Cattell (1995, S. 16) im Kontext des Zusammenhangs zwischen Intelligenz und Persönlichkeit darauf hin: „While we acknowledge the principle of parsimony and endorse it whenever applicable, the evidence points to relative complexity rather than simplicity. Insistence on parsimony at all costs can lead to bad science“. Modelle, in denen der Zusammenhang zwischen Intelligenz und Persönlichkeit abgebildet wird, unterliegen möglicherweise der Gefahr einer übermäßigen Vereinfachung. Zugleich stellt sich die Herausforderung der empirischen Überprüfbarkeit komplexer Modellannahmen. Im Folgenden werden Modelle mit unterschiedlicher Komplexität vorgestellt, die das Zusammenwirken zwischen Intelligenz und Persönlichkeit beschreiben (Ackerman & Heggestad, 1997; Chamorro-Premuzic & Arteche, 2008; Chamorro-Premuzic & Furnham, 2004).

#### *Das Modell von Ackerman und Heggestad (1997):*

Ackerman und Heggestad (1997) stellen ein integratives Modell vor, in dem ein mögliches Zusammenwirken von Intelligenz und Persönlichkeit unter Berücksichtigung des „Typical Intellectual Engagement“ (Typisches Intellektuelles Engagement; TIE; Ackerman, 1994; Goff & Ackerman, 1992) und der Interessen (Holland, 1963) dargestellt wird. In einer im Rahmen der Modellbildung durchgeführten Metaanalyse zum Zusammenhang zwischen Intelligenz und Persönlichkeit wurden 135 Studien inkludiert, bei denen unter Berücksichtigung der Three-Stratum Theorie (TST; Carroll, 1993) und den Persönlichkeitsinventaren zugrunde liegenden Theorien von Eysenck (1970), Tellegen (1982), Costa und McCrae (1992) sowie Digman (1990) von Ackerman und Heggestad (1997) eine Zuordnung zu den Fähigkeitskonstrukten bzw. Persönlichkeitsdimensionen erfolgte. Neben der metaanalytischen Untersuchung des Zusammenhangs zwischen Intelligenz und Persönlichkeit wurden Studien, in denen ein

empirischer Zusammenhang zwischen Intelligenz und TIE berichtet wurde, mit einer problematisch geringen Anzahl inkludierter Studien von eins bis sechs, ebenfalls einer Metaanalyse unterzogen. Des Weiteren erfolgten in der Arbeit von Ackerman und Heggestad (1997) Reviews zum Zusammenhang zwischen Interessen und Intelligenz mit fünf Studien sowie zum Zusammenhang zwischen Interessen und Persönlichkeit mit drei Studien. Unter Berücksichtigung der metaanalytischen Befunde und der Reviews wird von Ackerman und Heggestad (1997) ein Modell vorgeschlagen (Abbildung 4).



**Abbildung 4.** Intelligenzkonstrukte (fett), Interessen (normal) und Persönlichkeitsdimensionen (kursiv); durchgehende Linien: positive Korrelationen; gestrichelte Linien: negative Korrelationen. TIE = Typical Intellectual Engagement; nach Ackerman und Heggestad (1997, S. 238).

Unter Berücksichtigung der TST (Carroll, 1993) werden in Ackerman und Heggestad (1997) ausgewählte Persönlichkeitsdimensionen, Interessensbereiche und TIE schematisch nach der Zugehörigkeit zu den jeweiligen Intelligenzkonstrukten abgebildet. Überdies werden positive und negative Beziehungen zwischen den einzelnen Konstrukten durch unterschiedliche Linienarten gekennzeichnet. Als Grund, warum nicht alle substantiellen metaanalytischen Befunde in dem Modell berücksichtigt werden können, wird von den Autoren eine mangelnde Kohärenz ihrer metaanalytischen Befunde und der Literatur angenommen. Dennoch kann sich die Frage stellen, warum substantielle Populationszusammenhänge mit einer beachtlichen Anzahl inkludierter Studien ( $k = 30$  bis  $k = 66$ ), bspw. zwischen Intelligenz und Neurotizismus, nicht in das Modell integriert wurden. Denn es kann angenommen werden, dass es auch vor der Veröffentlichung von Ackerman und Heggestad (1997) kohärente Befunde und Theorien zum Zusammenhang von

Intelligenz und Neurotizismus gab (bspw. Most & Zeidner, 1995). Dessen ungeachtet wird eine Berücksichtigung interessenspezifischer Dimensionen aufgrund von Reviews mit fünf bzw. drei Studien im Modell zugelassen.

Ähnlich gering ist die Anzahl inkludierter Studien, die im Kontext des Konzeptes von TIE im Modell berücksichtigt wurden. Allerdings sollte angemerkt werden, dass Ackerman und Heggestad (1997) TIE als wichtiges theoretisches Konzept postulieren, was über den Zusammenhang metaanalytischer Befunde hinaus Varianz zwischen Intelligenz und Persönlichkeit erklären soll. Die Grundlage für diese Annahme bilden die Arbeiten von Goff und Ackerman (1992) sowie Ackermann (1994). Bedingt lässt sich damit erklären, weshalb TIE im Modell trotz der geringen Anzahl von Studien berücksichtigt wurde. Dessen ungeachtet ist die konzeptionelle Nähe zwischen bzw. die Unterscheidung von TIE und Offenheit für Erfahrungen Gegenstand häufiger Diskussionen (bspw. Arceche, Chamorro-Premuzic, Ackerman & Furnham, 2009; Chamorro-Premuzic & Furnham, 2004).

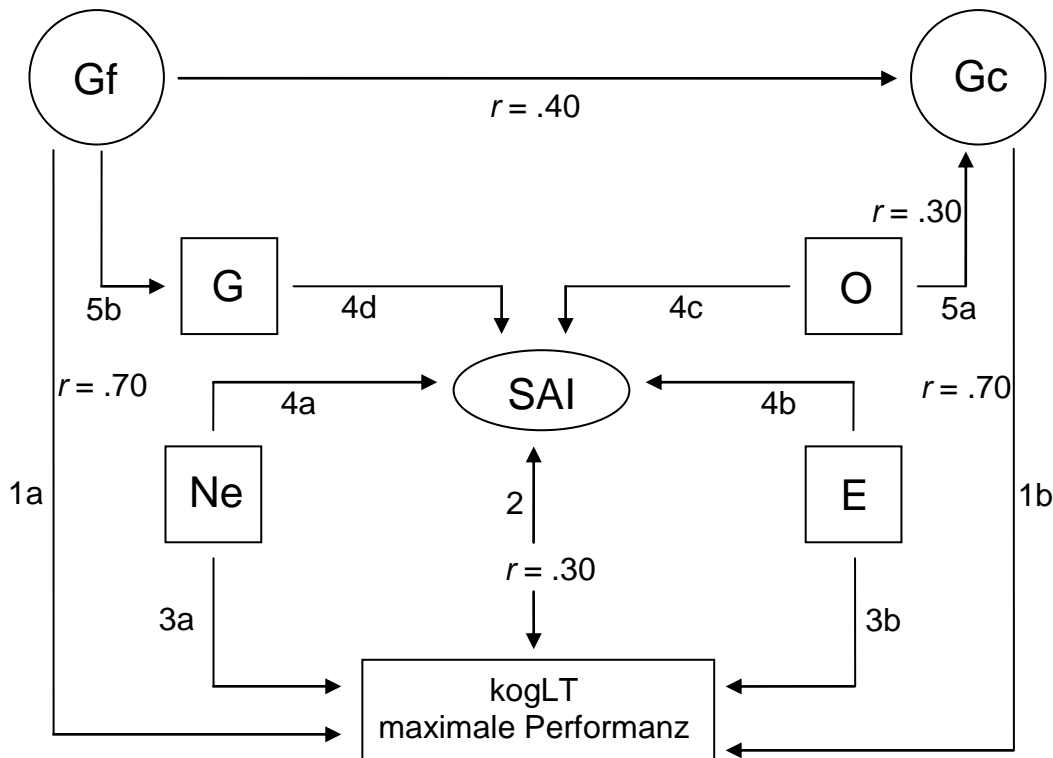
*Das Modell von Chamorro-Premuzic und Furnham (2004):*

In Chamorro-Premuzic und Furnham (2004) werden drei unterschiedliche Ebenen der Intelligenz angenommen:

- Die maximale kognitive Performanz, die mit kognitiven Leistungstests operationalisiert werden kann,
- die tatsächlichen intellektuellen Fähigkeiten, die nicht direkt operationalisierbar sind und
- die subjektiv selbsteingeschätzte Intelligenz (SAI), die mit Fragebögen operationalisiert werden kann.

Neben den theoretischen Annahmen zur Intelligenz wurde im Kontext der Persönlichkeit das FFM berücksichtigt. Die Korrelationskoeffizienten sind im Modell approximiert und basieren auf einer Stichprobengröße von  $N = 200$  (Chamorro-Premuzic & Furnham, 2004). Die Autoren berichten darüber hinaus, dass Korrelationskoeffizienten, die nicht im Modell berichtet sind, eine Korrelationshöhe von etwa .20 haben (Abbildung 5).





**Abbildung 5.** Zusammenhang zwischen Intelligenz und Persönlichkeit nach Chamorro-Premuzic und Furnham (2004, S. 253); Gf = fluide Intelligenz, Gc = kristallisierte Intelligenz; G = Gewissenhaftigkeit; O = Offenheit für Erfahrungen; SAI = subjektiv selbsteingeschätzte Intelligenz; Ne = Neurotizismus; E = Extraversion; kogLT = kognitiver Leistungstest; Pfeile ohne Korrelationshöhe  $r = .20$ ; Verträglichkeit nicht berücksichtigt ( $r < .20$ ).

Die Pfeile 1a und 1b bilden die Determination der maximalen kognitiven Performanz durch die tatsächlichen intellektuellen Fähigkeiten ab (Abbildung 5). Die maximale kognitive Performanz, messbar mit kognitiven Leistungstests, soll ein reliabler Indikator der tatsächlichen Intelligenz sein (Chamorro-Premuzic & Furnham, 2004). Der Doppelpfeil 2 (Abbildung 5) zeigt einen positiven Zusammenhang zwischen SAI und der maximalen kognitiven Performanz. Dabei wird bspw. von Chamorro-Premuzic und Furnham (2004) berichtet, dass Erfahrungen in der Bearbeitung von kognitiven Leistungstests dazu führen sollen, die eigene Intelligenz präziser einzuschätzen. Darüber hinaus berichten die Autoren, dass SAI ein besserer Prädiktor der maximalen kognitiven Performanz sein soll als die selbsteingeschätzte Persönlichkeitsdimension Offenheit für Erfahrungen. Begründet wird dies, trotz der konzeptionellen Überschneidung der beiden Konzepte, durch eine höhere Spezifität von SAI. Die Persönlichkeitsdimension Offenheit für Erfahrungen erfasst neben mehr intellektuellen Aspekten auch Aspekte über das Normen- und Wertesystem oder Aspekte über die Ästhetik einer Person (Chamorro-Premuzic & Furnham, 2004).

Chamorro-Premuzic und Furnham (2004) berichten, dass 30% der Varianz in kognitiven Leistungstests und anderen Prüfungssituationen durch Persönlichkeitseigenschaften determiniert

sein sollen. Neurotizismus soll einen negativen Einfluss auf die Ergebnisse in kognitiven Leistungstests haben (Pfeil 3a, Abbildung 5). Dabei wird angenommen, dass Individuen mit einer höheren Neurotizismusausrprägung ängstlicher und nervöser in Testsituationen reagieren, was letztlich die Performanz in kognitiven Leistungstests mindern soll (bspw. Moutafi, Furnham & Tsaousis, 2006). Hingegen soll sich eine höhere Extraversionsausprägung positiv auf die Performanz in kognitiven Leistungstests mit zeitlicher Begrenzung auswirken, da aufgrund eines optimaleren Erregungslevels die Bearbeitung aufmerksamer und reaktionsschneller erfolgen soll (Pfeil 3b, Abbildung 5; Eysenck, 1967; Rawlings & Carnie, 1989; Zeidner & Matthews, 2000). Allerdings gibt es hinsichtlich der kognitiven Leistungsfähigkeit in Abhängigkeit von der Extraversionsausprägung kontroverse Annahmen und Ergebnisse (Rawlings & Carnie, 1989; Rawlings & Skok, 1993; Robinson, 1985; Saklofske & Kostura, 1990). Darüber hinaus berichten neuere metaanalytische Untersuchungen zum Zusammenhang zwischen Intelligenz und Extraversion einen Populationseffekt von Null (Wolf & Ackerman, 2005).

Die Pfeile 4a, 4b, 4c und 4d (Abbildung 5) bilden die Annahme ab, dass die Persönlichkeitseigenschaften Neurotizismus, Extraversion, Offenheit für Erfahrungen und Gewissenhaftigkeit die Höhe der selbsteingeschätzten Intelligenz beeinflussen (Chamorro-Premuzic & Furnham, 2004). Personen mit einer höheren Neurotizismusausrprägung sollen dazu tendieren, ihre eigene Intelligenz zu unterschätzen. Hingegen sollen Personen mit einer höheren Extraversions- und/oder Offenheitsausprägung ihre Intelligenz überschätzen (Chamorro-Premuzic & Furnham, 2004).

Chamorro-Premuzic und Furnham (2004) nehmen darüber hinaus an, dass SAI eine Moderatorvariable zwischen der Persönlichkeit und der kognitiven Performanz ist. Fällt das Selbsturteil über die Intelligenz, bedingt durch eine höhere Extraversionsausprägung und geringere Neurotizismusausrprägung, positiver aus, soll dies zu einer optimistischeren Annäherung von zu bewältigenden Aufgaben und Problemen führen. Die kognitive Leistung soll in dem Fall nicht negativ beeinträchtigt sein. Entsprechend soll bei einer geringeren Extraversionsausprägung und höheren Neurotizismusausrprägung, das Selbsturteil über die Intelligenz geringer ausfallen, was zu einer Abnahme der kognitiven Leistung führen kann (Chamorro-Premuzic & Furnham, 2004; Zeidner & Matthews, 2000).

Die Pfeile 5a und 5b sollen das Zusammenwirken zwischen Persönlichkeit und den tatsächlichen intellektuellen Fähigkeiten abbilden. Dabei wird angenommen, dass eine höhere Offenheits-

## *Kapitel 2 – Theorie*

Ausprägung zu einer stärkeren Nutzung von Lerngelegenheiten und deshalb auch zu einer Zunahme von Gc führt (Pfeil 5a; siehe bspw. auch Ackerman, 1996; Ashton, Lee, Vernon & Jang, 2000). Mit dem Pfeil 5b soll die Determination der fluiden Intelligenz auf die Gewissenhaftigkeit abgebildet werden. Individuen mit einer geringeren Gf-Ausprägung versuchen durch Ordnung, Sorgfalt, Zielstrebigkeit und Beständigkeit die Nachteile in der fluiden kognitiven Leistungsfähigkeit zu kompensieren (Chamorro-Premuzic & Arteche, 2008; Chamorro-Premuzic & Furnham, 2004; Furnham, Moutafi & Chamorro-Premuzic, 2005; Moutafi, Furnham & Paltiel, 2004).

Chamorro-Premuzic und Furnham (2004) berücksichtigen in ihrem Modell die Investmenttheorie (Cattell, 1971, 1987), den Top-Down-Ansatz (Rindermann & Neubauer, 2001) sowie die Unterscheidung in die typische bzw. tatsächliche Intelligenz und die maximale kognitive Performanz (Cronbach, 1949; Eysenck & Eysenk, 1985). Die partielle empirische Bestätigung des Modells (Chamorro-Premuzic und Furnham, 2004) erfolgte von Chamorro-Premuzic und Arteche (2008).

### *Das Modell von Chamorro-Premuzic und Arteche (2008):*

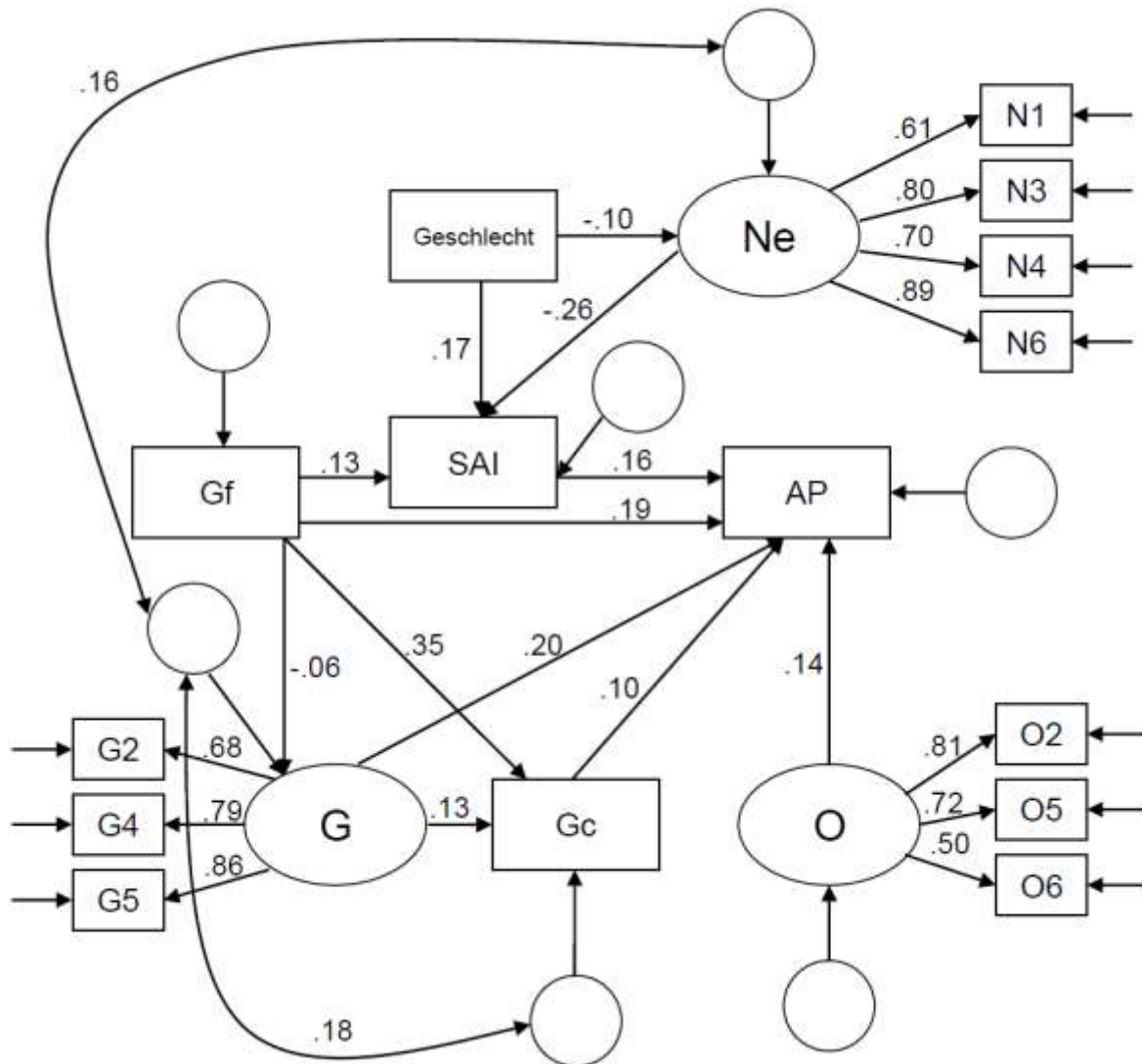
Das Modell aus Chamorro-Premuzic und Furnham (2004) wurde von Chamorro-Premuzic und Arteche (2008) modifiziert, um die theoretischen Modellannahmen über korrelative Beziehungen hinaus einer empirischen Überprüfung zu unterziehen. Über einen Zeitraum von vier Jahren wurden die Prüfungsleistungen von Studenten mit einer Stichprobengröße von  $N = 473$  in das Modell von Chamorro-Premuzic und Arteche (2008; Abbildung 6) integriert. Neben der Intention, mit Hilfe der Prüfungsleistungen (akademische Performanz; AP) Längsschnittdaten in das Modell zu integrieren, nehmen Chamorro-Premuzic und Arteche (2008) an, dass sich die AP nicht auf die maximale kognitive Performanz beschränkt, sondern darüber hinaus einen umfassenderen Ausschnitt einer Person hinsichtlich ihres Erfolges beinhaltet.

Es ist davon auszugehen, dass Restriktionen von den Autoren vorgenommen wurden, um das Modell mittels Strukturanalysen empirisch zu bestätigen. So werden letztlich folgende Subfacetten aus dem NEO-PI-R als beobachtete Variablen in dem Modell berücksichtigt:

- Ne: Ängstlichkeit (N1), Depression (N3), Soziale Befangenheit (N4) und Verletzlichkeit (N6),
- O: Offenheit für Ästhetik (O2), Offenheit für Ideen (O5), Offenheit für Werte- und Normensysteme (O6) und

- G: Ordnungsliebe (G2), Leistungsstreben (G4) und Selbstdisziplin (G5).

Als manifeste Variablen wurden Gf, Gc, SAI, AP und das Geschlecht berücksichtigt (siehe Abbildung 6).



**Abbildung 6.** Das Intelligenz-Persönlichkeitsmodell nach Chamorro-Premuzic und Arteche (2008, S. 566); Gf = fluide Intelligenz; Gc = kristalline Intelligenz; SAI = subjektiv selbsteingeschätzte Intelligenz; AP = akademische Perfomanz; Ne = Neurotizismus; N1 = Ängstlichkeit; N3 = Depression; N4 = soziale Befangenheit; N6 = Verletzlichkeit; O = Offenheit für Erfahrungen; O2 = Offenheit für Ästhetik (O2); O5 = Offenheit für Ideen, O6 = Offenheit für Werte- und Normensysteme; G = Gewissenhaftigkeit; G2 = Ordnungsliebe; G4 = Leistungsstreben; G5 = Selbstdisziplin; Fehlerterme wurden als leere Kreise dargestellt.

Das Modell von Chamorro-Premuzic und Arteche (2008) weist eine zufriedenstellende Anpassung an die Daten auf. Im Vergleich zu dem Modell von Chamorro-Premuzic und Furnham (2004), aber auch metaanalytischen Befunden aus Ackerman und Heggestad (1997) kann auffallen, dass die substantiellen Zusammenhänge zwischen Offenheit für Erfahrungen und der kristallisierten Intelligenz nicht abgebildet sind. Ebenso werden bzw. konnten wahrscheinlich bedeutsame

## *Kapitel 2 – Theorie*

Zusammenhänge zwischen Neurotizismus und der kristallisierten Intelligenz (Ackerman & Heggestad, 1997; Chamorro-Premuzic & Furnham, 2004) aufgrund einer empirischen Modellpassung nicht berücksichtigt werden.

### *Grenzen der empirischen Überprüfbarkeit der Modelle zum Zusammenwirken von Intelligenz und Persönlichkeit:*

In Anlehnung an die TST (Carroll, 1993) werden in Ackerman und Heggestad (1997) ausgewählte metaanalytische Befunde zum Zusammenhang von Intelligenz und Persönlichkeit in einem Modell integriert. Darüber hinaus wird das Konzept des TIE (Ackerman, 1994; Goff & Ackerman, 1992) und verschiedene Interessenbereiche (Holland, 1963) im Modell integriert. Es kann als positiv bewertet werden, dass eine Vielzahl unterschiedlicher Konzepte in dem Modell von Ackerman und Heggestad (1997) integriert sind. Zugleich können die Kriterien, weshalb ein Konzept im Modell berücksichtigt wurde, kritisiert werden. Denn die Inklusion von Einzelbefunden und die Nichtberücksichtigung metaanalytisch substantieller Befunde ist in dem Modell z.T. schwer nachvollziehbar.

Chamorro-Premuzic und Furnham (2004) berücksichtigen in ihrem Modell eine Vielzahl von Theorien (bspw. Cattell, 1971, 1987; Cronbach, 1949; Eysenck & Eysenck, 1985; Rindermann & Neubauer, 2001). Allerdings kann der Versuch, die theoretischen Annahmen mit selektierten korrelativen Maßen aus einer einzelnen Querschnittstudie ( $N = 200$ ) empirisch zu bestätigen, kritisiert werden. In dem Fall helfen die korrelativen Daten bestenfalls, Argumente plausibler zu machen (vgl. Wittmann, 1985).

Unter Berücksichtigung der AP (Längsschnittdaten), des Geschlechts, der selektiven Exklusion bzw. Inklusion verschiedener Subfacetten des NEO-PI-R, konnten Chamorro-Premuzic und Arteche (2008) das Modell von Chamorro-Premuzic und Furnham (2004) empirisch nachweisen. Dies ist durchaus als positiv zu bewerten. In einem folgenden Schritt könnten substantielle metaanalytische Befunde, bspw. zum Zusammenhang zwischen der kristallisierten Intelligenz und Offenheit für Erfahrungen, in dem Modell berücksichtigt werden.

Eine korrelative Beziehung ist eine notwendige, aber keine hinreichende Voraussetzung für eine kausale Abhängigkeit. Werden Modelle angestrebt, die kausale Annahmen berücksichtigen, sollten adäquate statistische Verfahren wie konfirmatorische Strukturgleichungsanalysen Anwendung finden. Neben der richtigen Wahl des statistischen Verfahrens sollte das Studiendesign gut

überlegt sein (siehe bspw. Querschnittsdesign vs. Längsschnittsdesign). Nicht zuletzt ist auch eine gute Datenqualität mit möglichst wenigen Artefakten, einschließlich die Berücksichtigung von Kontrollvariablen anzustreben. Diesbezüglich bieten sich metaanalytische Untersuchungen bzw. metaanalytische Befunde an, die über eine Vielzahl von Studien hinweg artefaktreduzierte Effektgrößen berichten. Im folgenden Kapitel 2.3.3 sollen die derzeit verfügbaren metaanalytischen Befunde zum Zusammenhang von Intelligenz und Persönlichkeit berichtet werden. Neben der Auflistung metaanalytischer Befunde bietet sich dabei die Möglichkeit einer vertiefenden Einordnung in die theoretischen Perspektiven an (bspw. Zeidner, 1995; Stumm et al., 2011). In dem Kapitel 2.3.4 werden Gründe aufgeführt, weshalb eine weitere Metaanalyse zum Zusammenhang von Intelligenz und Persönlichkeit sinnvoll sein kann.

### **2.3.3 Metaanalytische Ergebnisse zum Zusammenhang von g, Gf und Gc mit dem FFM aus vorherigen Studien**

*Ackerman und Heggestad (1997):*

Die Arbeit von Ackerman und Heggestad (1997) ist bisher eine der umfangreichsten Metaanalysen zum Zusammenhang von Intelligenz und Persönlichkeit. Zur Effektgrößenbestimmung wurde das „Fixed Effects Model“ (FE-Modell) nach Hedges und Olkin (1985) verwendet. Auf die unterschiedlichen Methoden zur Bestimmung von metaanalytischen Kennwerten wird in Kapitel 4.3 eingegangen.

In Tabelle 2 sind die Populationszusammenhänge von den Intelligenzfaktoren g, Gf, Gc mit den Persönlichkeitsdimensionen des FFM aus Ackerman und Heggestad (1997) abgebildet. Der Populationszusammenhang von g, Gf und Gc mit Neurotizismus (Ne) ist ausnahmslos kleiner null. Die Null liegt hierbei nicht im Konfidenzintervall (CI). Zwischen g, Gf, Gc und Extraversion (E) sowie g, Gc und Offenheit für Erfahrungen (O) ist ein Populationszusammenhang größer null nicht auszuschließen, da die Null nicht im CI liegt. Zum Zusammenhang zwischen Gf und O sowie g, Gf, Gc und Verträglichkeit (V) bzw. Gewissenhaftigkeit (G) kann ein Populationszusammenhang von null nicht ausgeschlossen werden, da in den Fällen die Null im CI liegt (Tabelle 2). Das durchschnittliche Mindestalter der Probanden aus den einzelnen inkludierten Studien beträgt 10 Jahre.

**Tabelle 2**

Zusammenhang zwischen g, Gf, Gc und den Persönlichkeitsdimensionen des FFM, zusammengestellt aus Ackerman und Heggestad (1997)

|              | $\rho$ | CI   | $k$ | $N$    |             | $P$ | CI   | $k$ | $N$    |
|--------------|--------|------|-----|--------|-------------|-----|------|-----|--------|
| <b>g/Ne</b>  | -.15   | -.12 | 30  | 6,169  | <b>g/E</b>  | .08 | .06  | 35  | 15,931 |
| <b>Gf/Ne</b> | -.08   | -.06 | 47  | 13,664 | <b>Gf/E</b> | .06 | .05  | 44  | 11,395 |
| <b>Gc/Ne</b> | -.09   | -.08 | 66  | 19,820 | <b>Gc/E</b> | .11 | .10  | 62  | 24,280 |
| <b>g/O</b>   | .33    | .26  | 3   | 555    | <b>g/V</b>  | .01 | -.06 | 6   | 941    |
| <b>Gf/O</b>  | .08    | -.06 | 2   | 205    | <b>Gf/V</b> | .03 | -.05 | 3   | 591    |
| <b>Gc/O</b>  | .30    | .25  | 10  | 1,695  | <b>Gc/V</b> | .04 | .00  | 10  | 2,206  |
| <b>g/G</b>   | .02    | -.01 | 3   | 4,850  |             |     |      |     |        |
| <b>Gf/G</b>  | -.08   | .02  | 4   | 401    |             |     |      |     |        |
| <b>Gc/G</b>  | -.05   | .01  | 7   | 1,485  |             |     |      |     |        |

*Anmerkungen.*  $\rho$  = Populationszusammenhang; CI = Konfidenzintervall, ausschließlich Angabe der Intervallgrenze, die zur Identifikation eines Nulleffektes benötigt wird;  $k$  = Anzahl der inkludierten Studien;  $N$  = Stichprobengröße; g/Ne = Zusammenhang zwischen allgemeiner Intelligenz und Neurotizismus; Gf/Ne = Zusammenhang zwischen fluider Intelligenz und Neurotizismus; Gc/Ne = Zusammenhang zwischen kristalliner Intelligenz und Neurotizismus; g/E = Zusammenhang zwischen allgemeiner Intelligenz und Extraversion; Gf/E = Zusammenhang zwischen fluider Intelligenz und Extraversion; Gc/E = Zusammenhang zwischen kristalliner Intelligenz und Extraversion; g/O = Zusammenhang zwischen allgemeiner Intelligenz und Offenheit für Erfahrungen; Gf/O = Zusammenhang zwischen fluider Intelligenz und Offenheit für Erfahrungen; Gc/O = Zusammenhang zwischen kristalliner Intelligenz und Offenheit für Erfahrungen; g/V = Zusammenhang zwischen allgemeiner Intelligenz und Verträglichkeit; Gf/V = Zusammenhang zwischen fluider Intelligenz und Verträglichkeit; Gc/V = Zusammenhang zwischen kristalliner Intelligenz und verträglichkeit; g/G = Zusammenhang zwischen allgemeiner Intelligenz und Gewissenhaftigkeit; Gf/G = Zusammenhang zwischen fluider Intelligenz und Gewissenhaftigkeit; Gc/G = Zusammenhang zwischen kristalliner Intelligenz und Offenheit für Erfahrungen.

*Die Metaanalyse von Wolf und Ackerman (2005):*

Die Metaanalyse von Wolf und Ackerman (2005) beschränkt sich auf die Untersuchung des Zusammenhangs zwischen Intelligenz und Extraversion. Das durchschnittliche Mindestalter der Probanden der einzelnen inkludierten Studien beträgt 10 Jahre. Zur Effektgrößenbestimmung wurde das FE-modell wie in Ackerman und Heggestad (1997) verwendet (siehe Kapitel 4.3). Im Vergleich zu Ackerman und Heggestad (1997) kann in Wolf und Ackerman (2005) eine Abnahme des positiven Populationszusammenhangs zwischen g, Gf, Gc und E festgestellt werden (Tabelle 3; Tabelle 2).

**Tabelle 3**

Der Zusammenhang von g, Gf, Gc mit E zusammengestellt aus Wolf und Ackerman (2005). Der inkludierte Studienzeitraum reicht von den Jahren 1961 bis 2003

|             | $\rho$ | CI  | <i>k</i> | <i>N</i> |
|-------------|--------|-----|----------|----------|
| <b>g/E</b>  | .05    | .04 | 70       | 24,688   |
| <b>Gf/E</b> | .05    | .03 | 70       | 16,617   |
| <b>Gc/E</b> | .08    | .07 | 103      | 33,844   |

*Anmerkungen.*  $\rho$  = Populationszusammenhang; CI = Konfidenzintervall, ausschließlich Angabe der Intervallgrenze, die zur Identifikation eines Nulleffektes benötigt wird; *k* = Anzahl der inkludierten Studien; *N* = Stichprobengröße; g/E = Zusammenhang zwischen allgemeiner Intelligenz und Extraversion; Gf/E = Zusammenhang zwischen fluider Intelligenz und Extraversion; Gc/E = Zusammenhang zwischen kristalliner Intelligenz und Extraversion.

In Tabelle 4 sind die Zusammenhänge zwischen Intelligenz, erfasst mit kognitiven Leistungstests, und Extraversion, erfasst mit selbstberichteten NEO-Inventaren, berichtet. Kann von einem negativen Populationszusammenhang in den Studien, in denen der NEO-FFI oder der NEO-PI zur Erfassung von E verwendet wurde, ausgegangen werden, sollte ein Nulleffekt in den Studien, in denen der NEO-PI-R verwendet wurde, nicht ausgeschlossen werden. Offenbar sind die konzeptionellen Unterschiede zwischen dem NEO-FFI oder NEO-PI und dem NEO-PI-R in Bezug auf Extraversion relevant für die Zusammenhänge mit Intelligenz.

**Tabelle 4**

Zusammenhang zwischen Intelligenz und Extraversion, getrennt nach NEO-FFI, NEO-PI und NEO-PI-R aus Wolf und Ackerman (2005)

| Moderator            | $\rho$ | CI   | <i>k</i> | <i>N</i> | Jahre <sup>1</sup> |
|----------------------|--------|------|----------|----------|--------------------|
| <b>NEO-Inventare</b> |        |      |          |          |                    |
| NEO-FFI              | -.10   | -.05 | 8        | 1,654    | 1997-2002          |
| NEO-PI               | -.08   | -.03 | 9        | 2,122    | 1992-2001          |
| NEO-PI-R             | .00    | -.04 | 10       | 2,498    | 1995-2003          |

*Anmerkungen.*  $\rho$  = Populationszusammenhang; CI = Konfidenzintervall, ausschließlich Angabe der Intervallgrenze, die zur Identifikation eines Nulleffektes benötigt wird; *k* = Anzahl der inkludierten Studien; *N* = Stichprobengröße; <sup>1</sup> = Inkludierter Studienzeitraum in Jahren; NEO-FFI = NEO-Five-Factor Inventory; NEO-PI = NEO-Personality Inventory; NEO-PI-R = NEO-Personality Inventory (revised); In Wolf und Ackerman (2005) keine Differenzierung von g, Gf und Gc innerhalb des Moderators NEO-Inventar.

*Die Metaanalyse von Judge, Jackson, Shaw, Scott und Rich (2007):*

In der Arbeit von Judge et al. (2007) wird der Zusammenhang zwischen der Selbstwirksamkeit und der beruflichen Leistung unter Berücksichtigung des FFM, der kognitiven Leistungsfähigkeit und der beruflichen Erfahrung metaanalytisch untersucht. Die für die vorliegende Arbeit relevanten Populationszusammenhänge zwischen der kognitiven Performanz und den Persönlichkeitsdimensionen des FFM sind in Tabelle 5 berichtet. Metaanalytische Kennwerte wie



das Glaubwürdigkeitsintervall (Credibility Interval; CRI), die Standardabweichung der Population ( $SD_\rho$ ) und die erklärte Varianz wurden in Judge et al. (2007) nicht berichtet. Judge et al. (2007) verwendete das Random-Effects-Model (RE-Modell) von Hunter und Schmidt (2004; siehe auch Kapitel 4.3).

**Tabelle 5**

Zusammenhänge zwischen der kognitiven Leistungsfähigkeit und den Persönlichkeitsdimensionen des FFM, zusammengestellt aus Judge et al. (2007) und Poropat (2009)

|           | Judge et al. (2007) |     |        | Poropat (2009) |     |        |
|-----------|---------------------|-----|--------|----------------|-----|--------|
|           | $\rho$              | $k$ | $N$    | $\rho$         | $k$ | $N$    |
| <b>Ne</b> | -.09                | 61  | 21,404 | -.06           | 47  | 31,955 |
| <b>E</b>  | .02                 | 61  | 21,602 | -.01           | 47  | 31,955 |
| <b>O</b>  | .22                 | 46  | 13,182 | .15            | 47  | 31,955 |
| <b>V</b>  | .00                 | 38  | 11,190 | -.03           | 47  | 31,955 |
| <b>G</b>  | -.04                | 56  | 15,429 | .01            | 47  | 31,955 |

Anmerkungen.  $\rho$  = wahre Effektgröße;  $k$  = Anzahl der inkludierten Studien;  $N$  = Stichprobengröße; Ne = Neurotizismus; E = Extraversion; O = Offenheit für Erfahrungen; V = Verträglichkeit; G = Gewissenhaftigkeit.

*Die Metaanalyse Poropat (2009):*

In der Arbeit von Poropat (2009) wird primär der Zusammenhang zwischen den Persönlichkeitsdimensionen des FFM und dem schulischen Erfolg metaanalytisch untersucht. Die für die vorliegende Arbeit relevanten Populationszusammenhänge zwischen der kognitiven Performanz und den Persönlichkeitsdimensionen des FFM sind in Tabelle 5 berichtet. Metaanalytische Kennwerte wie das CRI, die  $SD_\rho$  und die erklärte Varianz sind ebenfalls wie in Judge et al. (2007) nicht berichtet. Die Populationszusammenhänge aus Judge et al. (2007) und Poropat (2009) erlauben deshalb keine Interpretation hinsichtlich dessen, ob ein substantieller Effekt in der Population vorliegt bzw. ob ein Nulleffekt auszuschließen ist. Poropat (2009) verwendete das RE-Modell (siehe auch Kapitel 4.3). Das Durchschnittsalter der Probanden beträgt  $M = 17.21$  ( $SD = 3.77$ ).

*Vergleich bisheriger metaanalytischer Befunde zum Zusammenhang von Intelligenz und dem FFM:*

Zwischen der kognitiven Leistungsfähigkeit und der Neurozismusausprägung werden ausnahmslos negative Populationszusammenhänge berichtet, die zwischen  $\rho = -.06$  (Poropat, 2009) und  $\rho = -.15$  (Ackerman & Heggestad, 1997) variieren. Zwischen der kognitiven Leistungsfähigkeit und der Extraversionsausprägung sind die Populationszusammenhänge als inkonsistent über die verschiedenen metaanalytischen Untersuchungen zu bewerten. Zwischen der kognitiven Leistungsfähigkeit und der Offenheitsausprägung werden über die verschiedenen Metaanalysen hinweg positive Populationszusammenhänge mit  $\rho = .33$  (Ackerman & Heggestad, 1997),  $\rho = .22$

(Judge et al., 2007) und  $\rho = .15$  (Poropat, 2009) berichtet. Zum Zusammenhang von Verträglichkeit oder Gewissenhaftigkeit mit Intelligenz sollte ein Nulleffekt in der Population nicht ausgeschlossen werden.

#### **2.3.4 Betrachtungen zur Relevanz einer weiteren Metaanalyse zum Zusammenhang zwischen Intelligenz und Persönlichkeit**

Die Modelle aus Ackerman & Heggestad (1997), Chamorro-Premuzic und Furnham (2004) sowie Chamorro-Premuzic und Arteche (2008) lassen aufgrund der Berücksichtigung von g, Gf und Gc darauf schließen, dass in Bezug auf diese drei Fähigkeitskonstrukte die Konvergenzen erheblich sind. Wurden in Ackerman und Heggestad (1997) neben dem FFM noch Persönlichkeitskonzepte von Eysenck und Tellegen integriert, berücksichtigen neuere Arbeiten ausschließlich Persönlichkeitsdimensionen des FFM (bspw. Beauducel et al., 2007; Chamorro-Premuzic & Arteche, 2008; Chamorro-Premuzic & Furnham, 2004). Weitere Hinweise auf Konvergenzen innerhalb der Intelligenzmodelle und innerhalb der Persönlichkeitsmodelle finden sich bspw. in Ashton et al. (2000), Bates und Schieles (2003), Furnham und Chamorro-Premuzic (2006), Liepmann et al. (2008), McGrew (1997, 2005, 2009), Schulze und Roberts (2006), Süß (2001) sowie Süß und Beauducel (2011). Entsprechend dieser beobachteten Zunahme der Konvergenzen hinsichtlich der Fähigkeitskonstrukte g, Gf, Gc und der Persönlichkeitsdimensionen des FFM bietet sich nicht zuletzt eine weitere Metaanalyse an. Dabei ist davon auszugehen, dass konzeptionelle Zuordnungen verschiedener Aufgaben aus den kogLT zu den Fähigkeitskonstrukten sowie Zuordnungen verschiedener Persönlichkeitsdimensionen in das FFM homogener ausfallen bzw. weniger Subsummierungen bei der Aggregation vorgenommen werden müssen als in Ackerman und Heggestad (1997).

In Ackerman und Heggestad (1997) wurden die Beziehungen der Persönlichkeitsdimensionen aus unterschiedlichen Modellen (Costa & McCrae, 1992; Digman, 1990; Eysenck, 1970; Tellegen, 1982) konzeptionell zugeordnet. So wurden bspw. unter Extraversion die Persönlichkeitsdimensionen „Extraversion“, „Social Closeness“, „Social Potency“, „Well Being“ und „Positive Emotionality“ zusammengefasst. Dementsprechend kann bei einer Subsumierung verschiedener Konstrukte eine Heterogenität nicht ausgeschlossen werden. Dennoch wurde in Ackerman und Heggestad (1997) ein „fixed effects model“ (FE-Modell) zur Berechnung metaanalytischer Kennwerte verwendet, bei dem angenommen wird, dass die jeweiligen Konstrukte untereinander nicht variieren und aus einer gemeinsamen Population stammen. Die Folge einer fälschlicherweise angenommenen Homogenität kann zu einer Überschätzung des

Populationszusammenhangs führen (siehe auch Kapitel 4.3; Hedges & Olkin, 1985; Hunter & Schmidt, 2004; National Research Council, 1992). Dementsprechend kann eine Überschätzung des Populationszusammenhangs in Ackerman und Heggstad (1997) nicht ausgeschlossen werden.

Auch wenn mittlerweile von einer zunehmenden Konvergenz der Fähigkeitskonstrukte  $g$ ,  $G_f$  und  $G_c$  und einer zunehmenden Konvergenz der Persönlichkeitsdimensionen  $N_e$ ,  $E$ ,  $O$ ,  $V$  und  $G$  auszugehen ist, sollten verbleibende mögliche Unterschiede in den Studienpopulationsparametern nicht ausgeschlossen werden. Deshalb wurde in der vorliegenden Arbeit ein „random effects model“ (RE-Modell) zur Berechnung metaanalytischer Kennwerte verwendet. Ein RE-Modell hat die Möglichkeit, Variationen der Populationsparameter in den einzelnen Studien zuzulassen (siehe auch Kapitel 4.3). Durch die differenziertere Berücksichtigung einzelner Studien sollte sich die Genauigkeit und damit die Aussagekraft der metaanalytischen Befunde in der vorliegenden Arbeit erhöhen.

Die Arbeit von Ackerman und Heggstad (1997) stellt bis zum jetzigen Zeitpunkt die umfassendste metaanalytische Untersuchung zum Zusammenhang zwischen Intelligenz und Persönlichkeit dar. Allerdings ist die Anzahl inkludierter Studien, die den Zusammenhang zwischen  $g$ ,  $G_f$ ,  $G_c$  und einzelnen Dimensionen des FFM untersuchen, insgesamt als gering zu bewerten. Zum Zusammenhang von  $g$ ,  $G_f$  und  $G_c$  mit den Persönlichkeitsdimensionen  $O$ ,  $V$  und  $G$  variiert die inkludierte Stichprobenanzahl lediglich zwischen zwei und 10 Studien. Es kann davon ausgegangen werden, dass sich durch eine Vielzahl neuer Studien, die nach dem Jahr 1995 publiziert wurden, die Robustheit metaanalytischer Kennwerte zum Zusammenhang von  $g$ ,  $G_f$  und  $G_c$  mit den Persönlichkeitsdimensionen des FFM erhöht. Nicht zuletzt könnte es interessant sein, die metaanalytischen Ergebnisse aus Ackerman und Heggstad (1997) mit denen der vorliegenden Arbeit zu vergleichen.

Weitere Metaanalysen wurden nach 1997 veröffentlicht, die u.a. den Zusammenhang zwischen Intelligenz und Persönlichkeit untersuchen (Judge et al., 2007; Poropat, 2009; Wolf & Ackerman, 2005). Allerdings ist in den Arbeiten von Judge et al. (2007) und Poropat (2009) lediglich die Höhe der Populationszusammenhänge berichtet. Weitere metaanalytische Kennwerte wären notwendig, um die Populationszusammenhänge hinsichtlich eines Nulleffektes einordnen zu können. Weiterhin sind in Judge et al. (2007) und Poropat (2009) keine Zusammenhänge von  $G_f$  und  $G_c$  mit den Persönlichkeitsdimensionen des FFM berichtet. Darüber hinaus ist davon auszugehen, dass verschiedene Fähigkeitskonstrukte in Judge et al. (2007) unter Intelligenz

subsumiert sind. Um eine zu allgemeine bzw. ungenaue Subsumierung hinsichtlich der Fähigkeitskonstrukte zu vermeiden, wurde in der vorliegenden Arbeit u.a. eine Zuordnung der Fähigkeitskonstrukte g, Gf und Gc gemäß der CHC-Theorie (McGrew, 2009) nach der Vorlage von Süß & Beauducel (2011) vorgenommen.

In der Arbeit von Wolf und Ackerman (2005) wird metanalytisch der Zusammenhang zwischen g, Gf, Gc und Extraversion untersucht. Darüber hinaus werden auch folgende Moderatoren untersucht: Alter, kognitiver Leistungstest und Persönlichkeitsinventar. Die Anzahl inkludierter Studien in den Moderatoranalysen beträgt zwischen fünf und 23. Da Wolf und Ackerman (2005) ein FE-Modell zur Berechnung statistischer Kennwerte verwendet haben, ist eine Überschätzung der Effektgrößen, insbesondere innerhalb der Moderatoranalysen, nicht auszuschließen: Um mit dem FE-Modell die Populationseffektgröße bestimmen zu können, wird die Homogenität der inkludierten Populationsparameter vorausgesetzt, was mit Hilfe des  $\text{Chi}^2$ -Tests überprüft wird. Allerdings zeigt der  $\text{Chi}^2$ -Test die Variation der Studienpopulationsparameter bei einer inkludierten Studienanzahl kleiner als 30 nur unzureichend an (siehe Kapitel 4.3). Die Folge ist eine Überschätzung des Populationszusammenhangs (Hedges & Olkin, 1985; Hunter & Schmidt, 2004; National Research Council, 1992). Da in den Moderatoranalysen der vorliegenden Arbeit die Anzahl inkludierter Studien z.T. kleiner als 30 ist, wurde entsprechend das RE-Modell zur Bestimmung der metaanalytischen Kennwerte verwendet.

Das erste „Big-Five“-Inventar, der NEO-PI, wurde im Jahr 1985 von Costa und McCrae publiziert. Mit Beginn der NEO-Inventare ab dem Jahr 1985 lässt sich eine zunehmende Akzeptanz bis hin zu einer Dominanz des FFM in der Gegenwart beobachten (Schulze & Roberts, 2006). Aus diesem Grund wurde in der vorliegenden Arbeit das Jahr 1985 als die untere Inklusionsgrenze der zu interessierenden Studien festgelegt. Der häufige Einsatz verschiedener Persönlichkeitsinventare, denen neben den NEO-Inventaren ebenfalls das FFM zugrunde liegt, und die nach dem Jahr 1985 publiziert wurden, bspw. der International Personality Item Pool (Goldberg, 1992), das Personal Characteristics Inventory (Barrick & Mount, 1995), der Vijf Persoonlijkheids-Factoren Test (Elshout & Akkerman, 1975) oder das Perspectives Inventory (Abraham & Morrison, 2002), erlaubt darüber hinaus die Untersuchung konzeptioneller Variationen in der Operationalisierung innerhalb der Konstrukte des FFM mittels der RE-Methode.

Das Mindestalter der Probanden der inkludierten Studien beträgt in Ackerman und Heggestad (1997) sowie Wolf und Ackerman (2005) 10 Jahre. In einer Metaanalyse zur Stabilität von Persönlichkeit in Abhängigkeit vom Alter (Roberts & DelVecchio, 2000) wird eine Zunahme der

artefaktkorrigierten Retestrelabilität von .31 bei Kindern auf .54 bei Jugendlichen berichtet. Dieser Befund geht mit älteren Arbeiten einher, die die Stabilität der Persönlichkeit im Alter unter 15 Jahren als gering einschätzen (Eysenck & Cookson, 1969). Um eine Stabilität der Persönlichkeit hinsichtlich des Alters zu gewährleisten, wurde in der vorliegenden Arbeit das Mindestalter der Probanden von 10 Jahren (vgl. Ackerman & Heggstad, 1997; Wolf & Ackerman, 2005) auf 15 erhöht.

Oftmals wird in Metaanalysen eine Zuordnung der interessierenden Konstrukte aus den inkludierten Studien im Nachhinein durchgeführt (bspw. Ackerman & Heggstad, 1997; Judge et al., 2007). Im günstigsten Fall werden die Zuordnungen anhand bestehender Theorien bzw. Modelle vorgenommen. Bspw. wurden in Ackerman und Heggstad (1997) die Aufgabentypen kognitiver Leistungstests aus den inkludierten Studien gemäß der TST (Carroll, 1993) im Nachhinein vorgenommen. Insgesamt wurden von den Autoren 11 Kategorien gebildet: Die Allgemeine Intelligenz auf dem dritten Stratum, die acht Fähigkeitskonstrukte des zweiten Stratum (siehe Abbildung 1, S. 5) und zwei weitere Fähigkeitskonstrukte auf dem ersten Stratum („Math Reasoning, Closure“). Überdies wurde eine Kategorie „Miscellaneous Ability“ gebildet. Diese letztgenannte Kategorie sollte für die Fähigkeitskonstrukte bestimmt sein, die nicht den anderen 10 Fähigkeitskonstrukten zuzuordnen waren. Die Zuordnung der Fähigkeitskonstrukte aus den inkludierten Studien zu den 11 Kategorien führten Ackerman und Heggstad (1997) getrennt durch. Die Übereinstimmung der Zuordnung betrug 96%. Die Übereinstimmung der getrennten Zuordnung der verschiedenen Persönlichkeitskonstrukte von Ackerman und Heggstad (1997) war mit 61% bis 79% deutlich geringer. Insbesondere unter dem Aspekt der geringen Interrater-Übereinstimmung ist es schwer nachvollziehbar, warum die Autoren das FE-Modell zur Bestimmung metaanalytischer Kennwerte verwendet haben, bei dem eine Homogenität in der Population vorausgesetzt wird (siehe oben; siehe Kapitel 4.3).

Um eine Einschätzung der Auswirkungen unterschiedlicher Zuordnungen der Fähigkeitskonstrukte auf die metaanalytischen Kennwerte zu untersuchen, wurden zwei Analysen durchgeführt. In der ersten metaanalytischen Untersuchung wurde die Klassifizierung der Fähigkeitskonstrukte, wie sie von den Autoren der publizierten Studien erfolgte, beibehalten. Dadurch ist die Variabilität in der Zuordnung von Aufgaben zu Fähigkeitskonstrukten in der Variabilität der Effektstärken enthalten. In der zweiten Analyse-Variante wurde eine nachträgliche Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) nach der Zuordnungsvorlage von Süß und Beauducel (2011) durch den Autor der vorliegenden Arbeit vorgenommen. Wie oben bereits

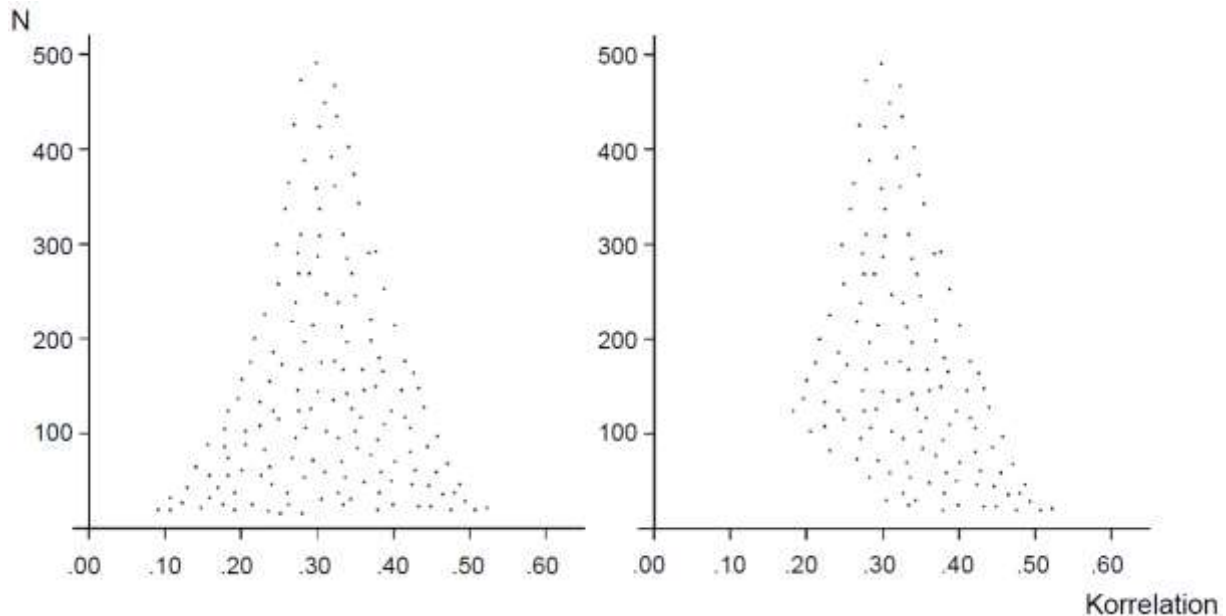
erwähnt, sollen die beiden Analyse-Varianten einen Vergleich hinsichtlich der Auswirkungen unterschiedlicher Zuordnungen der Fähigkeitskonstrukte auf die metaanalytischen Kennwerte liefern.

Eine Kritik, die häufig im Kontext von Metaanalysen geäußert wird, ist eine Ungenauigkeit metaanalytischer Kennwerte aufgrund der Nichtverfügbarkeit aller existierenden Studien zu einer Thematik (bspw. Hunter & Schmidt, 2004; Rothstein & Bushman, 2012). Gelangen Studien, die signifikante Effektgrößen berichten, häufiger zur Publikation in wissenschaftlichen Fachzeitschriften als nichtsignifikante Effektgrößen, entsteht eine Veröffentlichungsverzerrung. Eine Veröffentlichungsverzerrung bzw. ein sogenanntes Publikationsbias würde in dem Fall zu einer Überschätzung der wahren Effektgröße führen (Begg, 1994; Coursol & Wagner, 1986; Hunter & Schmidt, 2004). Bisher wurde jedoch das Publikationsbias im metaanalytischen Kontext zum Zusammenhang von Intelligenz und Persönlichkeit nicht untersucht.

### **2.3 Publikationsbias**

Das Publizieren in wissenschaftlichen Fachzeitschriften ist u.a. in der Psychologie unvermeidbar, um erfolgreich eine wissenschaftliche Karriere zu durchlaufen. Dabei kann angenommen werden, dass neue und innovative Forschungsthemen in Fachzeitschriften im Gegensatz zu Replikationsarbeiten vermehrt publiziert werden, denen darüber hinaus nichtsignifikante Effektgrößen zugrunde liegen (bspw. Coursol & Wagner, 1986; Ferguson & Heene, 2012). Die Tendenz, Replikationsarbeiten zu vermeiden, und/oder signifikante Effektgrößen bevorzugt zu publizieren, führt zu einem Publikationsbias und letztlich zu einer eingeschränkten Glaubwürdigkeit der Forschung (Ferguson & Heene, 2012). Obwohl das Thema Publikationsbias kontrovers behandelt wird, besteht ein reges Streben, Effektgrößen zu einer Thematik hinsichtlich eines Publikationsbias zu untersuchen (bspw. Coursol & Wagner, 1986; Duval & Tweedie, 2000a, 2000b; Ferguson & Brannick, 2012; Ferguson & Heene, 2012; Hedges & Olkin, 1985; Hunter & Schmidt, 2004; Light & Pillemer, 1984; Rosenthal, 1979; Rothstein & Bushman, 2012). Zur Identifikation des Publikationsbias wurden verschiedene Verfahren entwickelt, die zumeist entweder auf der Annahme beruhen, dass vermehrt Studien mit signifikanten Effektgrößen publiziert werden (bspw. Hedges, 1992; Hedges & Olkin, 1985; Iyengar & Greenhouse, 1988; Rosenthal, 1979) oder auf der Annahme beruhen, dass bei einer vollständigen Inklusion aller Effektgrößen zu einer Thematik eine symmetrische Verteilung, grafisch ähnlich einer Normalverteilungskurve oder einem umgekehrten Trichter („funnel“), zu erwarten ist (Abbildung

7, links). Die Effektgrößen, denen eine kleinere Stichprobengröße zugrunde liegt, unterliegen größeren Stichprobenfehlern und weisen deshalb eine größere Streuung auf. Mit Zunahme der Stichprobengröße verringert sich der Stichprobenfehler und damit auch die Streuung. Fluktuieren einzelne Effektgrößen auffällig, was mit einer asymmetrischen Verteilung einhergeht, sollte ein Publikationsbias nicht ausgeschlossen werden (bspw. Begg & Mazumdar, 1994; Duval & Tweedie, 2000a, 2000b; Light & Pillemer, 1984; Abbildung 7, rechts).



**Abbildung 7.** Beispiel eines symmetrischen (links) und asymmetrischen (rechts) Streudiagramms (fiktiv); N = Stichprobengröße.

Im Rahmen der vorliegenden Arbeit soll das Publikationsbias zum Zusammenhang zwischen Intelligenz und Persönlichkeit mit zwei unterschiedlichen Methoden untersucht werden. Die symmetrische Verteilung publizierter Effektgrößen berücksichtigend, soll ein bereits etabliertes Verfahren von Duval und Tweedie (2000a, 2000b) eingesetzt werden, nach dem im Falle einer asymmetrischen Verteilung der publizierten Effektgrößen ein Publikationsbias nicht ausgeschlossen werden sollte. Das zweite Verfahren, welches im Rahmen der vorliegenden Arbeit entwickelt wurde, berücksichtigt, dass vermehrt Studien mit signifikanten Effektgrößen publiziert werden. Dazu soll mittels Simulation eine zufällige Verteilung signifikanter und nichtsignifikanter Effektgrößen – ohne Publikationsbias – erzeugt und mit dem Verhältnis der tatsächlich publizierten signifikanten und nichtsignifikanischen Effektgrößen – mit möglichem Publikationsbias – verglichen werden. Im Methodenteil unter 4.6 (Spezifikation verwendeter Methoden zur Identifikation des Publikationsbias) wird auf die zwei in der vorliegenden Arbeit verwendeten Verfahren im Detail eingegangen.

### **3 Zielsetzungen**

Das Ziel der vorliegenden Arbeit ist es, mittels RE-Methode, den Zusammenhang von den Fähigkeitskonstrukten – Allgemeine Intelligenz (g), Fluide Intelligenz (Gf), Kristallisierte Intelligenz (Gc) – mit den Persönlichkeitsdimensionen – Neurotizismus (Ne), Extraversion (E), Offenheit für Erfahrung (O), Verträglichkeit (V) und Gewissenhaftigkeit (G) – metaanalytisch zu untersuchen. Dabei wird von der Einschätzung ausgegangen, dass die Autoren der publizierten Studien eine z.T. liberale bzw. ungenaue Zuordnung der jeweiligen Intelligenztests bzw. kognitiven Leistungsaufgaben zu den Fähigkeitskonstrukten g, Gf oder Gc vorgenommen haben. Um diesbezügliche Ungenauigkeiten zu untersuchen, sollen zwei Analyse-Varianten durchgeführt werden. In der ersten metaanalytischen Analyse-Variante soll die Klassifizierung der Fähigkeitskonstrukte, wie sie von den Autoren der publizierten Studien erfolgte, beibehalten werden. Dadurch soll die Variabilität in der Zuordnung von Aufgaben zu Fähigkeitskonstrukten in der Variabilität der Effektstärken enthalten sein. In einer zweiten Analyse-Variante soll die Klassifizierung der Fähigkeitskonstrukte nach einer einheitlichen Intelligenzstrukturtheorie, der CHC-Theorie, im Nachhinein durch den Autor der vorliegenden Arbeit vorgenommen werden. Als Zuordnungsvorlage sollen hierbei die Arbeiten von McGrew (1997, 2005, 2009) und Süß und Beauducel (2011) berücksichtigt werden. Durch den Vergleich der beiden Analyse-Varianten bietet sich die Möglichkeit, Auswirkungen unterschiedlicher Zuordnungen der Fähigkeitskonstrukte auf die metaanalytischen Kennwerte einzuschätzen.

Darüber hinaus soll der Einfluss der Studien mit sehr großen Stichproben auf die metaanalytischen Kennwerte untersucht werden (vgl. Hunter & Schmidt, 2004). Dazu sollen zunächst die Studien mit sehr großen Stichproben inkludiert bleiben. Anschließend, in einer weiteren Analyse, sollen diese Studien exkludiert werden. Dadurch soll eingeschätzt werden, wie groß der Einfluss von Studien mit sehr großen Stichproben auf die metaanalytischen Kennwerte ist. Außerdem soll untersucht werden, inwieweit die metaanalytischen Ergebnisse aus Ackerman und Heggestad (1997), Judge et al. (2007), Poropat (2009) sowie Wolf und Ackermann (2005) mit den metaanalytischen Ergebnissen aus der vorliegenden Arbeit replizierbar sind.

In den Fällen, in denen die erklärte Varianz des Populationszusammenhangs kleiner als 75% sind, sollen mittels Moderatoranalysen Subpopulationen identifiziert werden (vgl. Hunter & Schmidt, 2004; Kemery, Mossholder & Roth, 1987). Dazu sollen verschiedenen kogLt und Persönlichkeitsinventare als Moderatoren berücksichtigt werden (siehe Kapitel 4.5).



### *Kapitel 3 – Zielsetzungen*

Ein weiteres zentrales Anliegen der vorliegenden Arbeit ist es, verschiedene Aspekte des Publikationsbias zum Zusammenhang von  $g$ ,  $G_f$  und  $G_c$  mit den Persönlichkeitsdimensionen des FFM zu untersuchen. Dem Vorschlag von Rothstein und Bushman (2012) folgend, möglichst verschiedene Methoden zur Identifikation des Publikationsbias zu nutzen, sollen zwei verschiedene Verfahren berücksichtigt werden. Mit der bereits publizierten und etablierten „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) soll die symmetrische Verteilung der publizierten Effektgrößen untersucht werden. Falls die publizierten Effektgrößen nicht symmetrisch um die mittlere Effektgröße verteilt sind, kann nach Duval und Tweedie (2000a, 2000b) oder Duval (2005) ein Publikationsbias nicht ausgeschlossen werden. Um nicht ausschließlich von einer Annahme auf ein Publikationsbias zu schließen, soll eine weitere, im Rahmen der vorliegenden Arbeit entwickelte Methode Anwendung finden: Mit Hilfe von Simulationen soll eine zufällige Verteilung signifikanter und nichtsignifikanter Effektgrößen um den Populationszusammenhang – ohne Publikationsbias – erzeugt werden. Anschließend soll die simulierte Verteilung mit der Verteilung der tatsächlich publizierten signifikanten und nichtsignifikanten Effektgrößen – mit möglichen Publikationsbias – verglichen werden. Tritt eine Differenz zwischen der simulierten Verteilung als Referenz und der empirischen Verteilung auf, sollte ein Publikationsbias nicht ausgeschlossen werden. Bei Annahme eines Publikationsbias nach der Simulationsmethode soll überdies die Anzahl der zusätzlichen Studien berechnet werden, die notwendig wären, um ein Publikationsbias ausschließen zu können (siehe Kapitel 4.6).

## 4 Methode

### 4.1 Literaturrecherche

Für die Literaturrecherche wurde die Online-Datenbank „PsycINFO“ verwendet. Anhand folgender Suchbegriffe wurden Studien in einem Veröffentlichungszeitraum der Jahre von 1985 bis 2011 gesucht:

- **Intelligenz:** "intelligence" oder "IQ" oder "mental ability" oder "cognitive ability" oder "general mental ability" oder "fluid intelligence" oder "crystallized intelligence" oder "g" oder "gf" oder "gc" **und**
- **Persönlichkeit:** "neuroticism" oder „emotional stability“ oder "extraversion" oder „openness“ oder "conscientiousness" oder "agreeableness“.

Darüber hinaus wurden die Literaturverzeichnisse publizierter Metaanalysen, die den Zusammenhang zwischen Intelligenz und Persönlichkeit untersuchen, gesichtet (Ackerman & Heggestad, 1997; Judge et al., 2007; Poropat, 2009; Wolf & Ackerman, 2005).

### 4.2 Studieninklusion und Zuordnung

Studien wurden inkludiert, wenn folgende Voraussetzungen gegeben waren:

- die Probanden Menschen sind,
- die Sprache Englisch ist,
- die Studien zwischen den Jahren 1985 und 2011 publiziert wurden,
- der Altersdurchschnitt der Probanden der jeweiligen Stichprobe  $M_{\text{Alter}} - SD_{\text{Alter}} \geq 15$  Jahre ist,
- die Stichprobe keine klinischen Störungen hat, um generelle Aussagen zur Population machen zu können,
- die korrelativen Zusammenhänge zwischen den Ergebnissen eines kognitiven Leistungstests und den Ergebnissen eines selbstberichteten Persönlichkeitsinventars anhand einer Stichprobe berichtet sind,
- die korrelativen Zusammenhänge zwischen mindestens einem der kognitiven Fähigkeitskonstrukte g, Gf, Gc und mindestens eine der Persönlichkeitsdimension Ne, E, O, V, G explizit in der Studie berichtet sind oder
- eine Zuordnung zu den Fähigkeitskonstrukten g, Gf, Gc nach der CHC-Theorie (McGrew, 1997, 2005, 2009) möglich war (siehe unten) und der korrelative Zusammenhang zwischen mindestens einem der nach der CHC-Theorie zugeordneten Fähigkeitskonstrukte und mindestens einer der Persönlichkeitsdimensionen Ne, E, O, V, G berichtet ist.

**Tabelle 6**

Sukzessiver Selektionsprozess geeigneter Studien zur Durchführung einer Metaanalyse in der vorliegenden Arbeit

|  | Initialsuche <sup>1</sup> | relevante Zusammenfassungen <sup>2</sup> | kodierte Datensätze gemäß Autoren publizierter Studien <sup>3</sup> |            |           | kodierte Datensätze nach CHC <sup>4</sup> |            |            |
|--|---------------------------|--|---|------------|-----------|---|------------|------------|
|  |                           |  | <b>g</b>  | <b>Gf</b>  | <b>Gc</b> | <b>g</b>                                  | <b>Gf</b>  | <b>Gc</b>  |
| g,Gf,Gc-Neurotizismus/-emotionale Stabilität | 641/104                   | 114                                      | 69  | 26         | 10        | 76  | 55         | 30         |
| g,Gf,Gc-Extraversion                         | 616                       | 92                                       | 71  | 26         | 12        | 86  | 60         | 33         |
| g,Gf,Gc-Offenheit                            | 509                       | 99                                       | 70  | 28         | 13        | 74  | 51         | 34         |
| g,Gf,Gc-Verträglichkeit                      | 244                       | 74                                       | 57  | 22         | 8         | 59  | 42         | 21         |
| g,Gf,Gc-Gewissenhaftigkeit                   | 466                       | 107                                      | 77  | 23         | 9         | 75  | 45         | 23         |
| <b>Σ</b>                                     | <b>2,580</b>              | <b>486</b>                               | <b>344</b>  | <b>125</b> | <b>52</b> | <b>370</b>                                | <b>253</b> | <b>141</b> |

Anmerkungen.<sup>1</sup> = Zeitpunkt der Recherche vom 15.11-23.11.2011; <sup>2</sup> = Anzahl der Studien, bei denen das Lesen des „Abstract“ nicht zur Exklusion geführt hat; <sup>3</sup> = inkludierte unabhängige Datensätze, in denen eine Zuordnung von g, Gf oder/und Gc explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>4</sup> = inkludierte unabhängige Datensätze, bei denen eine Zuordnung der Fähigkeitskonstrukte durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde; die numerischen Werte beziehen sich auf die Anzahl der Studien bzw. Datensätze.

In Tabelle 6 sind unter „Initialsuche“ die Anzahl der Studien aufgelistet, die bei der Eingabe der jeweiligen Kombinationen zwischen Intelligenz und Persönlichkeit (siehe oben) in der Online-Datenbank „PsycINFO“ angezeigt wurden und bei denen die Zusammenfassung („Abstract“) mindestens so weit gelesen wurde, bis ein Inklusionskriterium nicht erfüllt war. Wenn durch das Lesen des „Abstract“ kein Inklusionskriterium verletzt wurde, wurde die Studie weiter berücksichtigt (siehe Tabelle 6, relevante Zusammenfassungen). In den Spalten „kodierte Datensätze gemäß Autoren publizierter Studien“ und „kodierte Datensätze gemäß CHC“ ist die Anzahl der Studien berichtet, die den Inklusionskriterien entsprachen und letztlich kodiert werden konnten. Insgesamt konnten 154 Studien, denen 161 unabhängige Datensätze zugrundeliegen, berücksichtigt werden. Die Stichprobengröße beträgt  $N = 154,721$ , die durchschnittliche Stichprobengröße pro Studie  $\bar{n} = 961$  (Range: 38 bis 71,887,  $SD = 5,896.13$ ). In 133 Datensätzen (82.61%) können Aussagen über die Geschlechterverteilung gemacht werden. Dabei beträgt der Anteil weiblicher Probanden 42.25% und der Anteil männlicher Probanden 57.75%. Zwei der Stichproben bestehen ausschließlich aus weiblichen Teilnehmern (Guerin, et al., 2011; Staudinger, Maciel, Smith & Baltes, 1998) und zwei Stichproben bestehen ausschließlich aus männlichen Teilnehmern (Fine, 2007; Wickett & Vernon; 2000). Für 121 Studien ist das durchschnittliche Alter der Teilnehmer berichtet, es liegt zwischen 15.48 und 79.10 Jahren ( $M = 27.00$ ,  $SD = 12.70$ ).

*Besonderheiten während des Inklusionsprozesses:*

In einigen Studien wurden Zusammenhangsmaße aus mehreren Datensätzen in einem Artikel berichtet (bspw. Furnham & Chamorro-Premuzic, 2006). Wenn den Datensätzen unterschiedliche Stichproben zugrunde liegen, wurden diese als unabhängige Studien kodiert. Dazu berichten Hunter und Schmidt (2004, S.440): „Statistically, outcome values for nonoverlapping groups have the same properties as values from different studies“. Wenn hingegen nicht ausgeschlossen werden konnte, dass gleiche Stichproben in unterschiedlichen referierten Fachzeitschriften verwendet wurden, wurde nur ein Datensatz inkludiert (siehe bspw. Austin, Deary & Gibson, 1997; Austin, Deary, & Willock, 2001)

Wenn unterschiedliche Fähigkeitskonstrukte von den Autoren der publizierten Studien berichtet wurden, wurde das spezifischere, in der vorliegenden Arbeit interessierende, Fähigkeitskonstrukt berücksichtigt. In De Fruyt, Aluja, Garcia, Rolland und Jung (2006) werden z. B. die “Standard Progressive Matrices” (SPM) als kognitiver Leistungstest eingesetzt. Es wird von den Autoren berichtet, dass die SPM sowohl Gf als auch g erfassen: „...considered as a basic measure of fluid intelligence and one of the best measures of g“ (De Fruyt et al., 2006, S.104). In dem Fall wurde das spezifischere Intelligenzkonstrukt – Gf – berücksichtigt.

Wurden Effektgrößen separat zwischen Frauen und Männern berichtet, wurde eine Mittelung unter Berücksichtigung von Fischer’s Z-Transformation (Fischer, 1918) durchgeführt (bspw. Carson, Stalikas & Bizot, 1997; Carless, 1999). Unabhängig davon wurden anhand unterschiedlicher IQ-Werte verschiedene Gruppen der Studienteilnehmer von den Autoren der veröffentlichten Studien gebildet (bspw. De Fruyt et al., 2006). Die Annahme berücksichtigend, dass sich die Ausprägung der Intelligenz entsprechend der Gauß-Funktion verteilt, wurden die Stichprobengrößen der jeweiligen IQ-Gruppen unter Berücksichtigung der Normalverteilungskurve gewichtet. Die so erzeugten Gewichtungen ( $n_{gew1}, n_{gew2}, n_{gewi} \dots$ ) wurden mit der den jeweiligen dazugehörigen Korrelation der IQ-Gruppe multipliziert ( $r_1, r_2, r_i \dots$ ), anschließend addiert und letztlich durch die gesamte Stichprobengröße der Studie ( $N_{Studie}$ ) dividiert (Formel 1). Die so berechnete mittlere Korrelationshöhe  $\bar{r}$  wurde weiter berücksichtigt.

$$\bar{r} = \frac{\sum_{i=1}^g r_i n_{gew_i}}{N_{Studie}}, g = \text{Anzahl der Gruppen} \quad (1)$$

Wenn mehrere Stichproben nicht eindeutig den berichteten Korrelationen innerhalb einer Studie zuzuordnen war (bspw. Arteche, Chamorro-Premuzic, Furnham & Crump, 2008), wurde konservativ die kleinste Stichprobengröße berücksichtigt. In Arteche et al. (2008) wurden Stichprobengrößen von  $n = 307$  bis  $n = 585$  berichtet. Da in diesem Fall keine eindeutige Zuordnung der Stichprobengröße zu den berichteten Korrelationen möglich war, wurde die kleinste Stichprobengröße von  $n = 307$  berücksichtigt.

*Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien vs. Zuordnung der Fähigkeitskonstrukte gemäß CHC-Theorie (McGrew, 1997, 2005, 2009):*

In einer ersten Analyse-Variante blieben die Fähigkeitskonstrukte gemäß der Autoren publizierter Studien unverändert. In einer zweiten Analyse-Variante wurde eine Zuordnung der Fähigkeitskonstrukte im Nachhinein durch den Autor der vorliegenden Arbeit gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen. Als Grundlage für die Zuordnung der Fähigkeitskonstrukte  $g$ ,  $Gf$  und  $Gc$  gemäß der CHC-Theorie diente die Arbeit von McGrew (2009) sowie Süß und Beauducel (2011). Im Folgenden sollen Beispiele aufgeführt werden, wie die Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde:

- Aufgrund ungenauer bzw. zu verallgemeinernden Zuordnungen spezifischer Fähigkeitskonstrukte zu  $g$  (bspw. Austin et al., 1997; Bipp, Steinmayr & Spinath, 2008; Deuling, Denissen, van Zalk, Meeus & van Aken, 2011; Egan, 1989; Freund & Holling, 2011, Furnham, Taylor & Chamorro-Premuzic, 2008; Miller & Tal, 2007; Yeo & Neal, 2008), wurden unter Berücksichtigung der CHC-Theorie (McGrew, 2009) Verschiebungen zu spezifischeren Fähigkeitskonstrukten durchgeführt. Wenn bspw. die Verwendung des „Raven’s Progressive Matrices“ als Marker für  $g$  berichtet wird (Yeo & Neal, 2008), erfolgte eine Zuordnung nach der CHC-Theorie, „reasoning“, zu  $Gf$  (vgl. McGrew, 2009).
- Wenn die Fähigkeitskonstrukte  $g$ ,  $Gf$  und  $Gc$  nicht explizit von den Autoren der publizierten Studien berichtet sind, wurde wenn es möglich war, eine Zuordnung gemäß CHC-Theorie (McGrew, 2009) vorgenommen (bspw. Allworth & Hesketh, 2000; Bartone, Snook & Tremble, 2002; Carson et al., 1997; Chamorro-Premuzic, Furnham & Ackerman, 2006a; Hultsch, Hertzog, Small & Dixon, 1999; Sperbeck, Whitbourne & Hoyer, 1986). Dazu ein Beispiel: In Hultsch et al. (1999) wurden u.a. die Fähigkeitskonstrukte „reading comprehension“ und „vocabulary“ als Teil einer kognitiven Testbatterie erfasst. In dem Fall wurde nach der Beschreibung von McGrew (2009) „reading comprehension“ zu  $Gf$  und „vocabulary“ zu  $Gc$  zugeordnet. Es kam darüber hinaus auch vor, dass keine explizite

Zuordnung zu einem spezifischen Fähigkeitskonstrukt durch die Autoren der publizierten Studie erfolgte. Dazu ein weiteres Beispiel: Bei der Verwendung des „General Knowledge Test“ (GKT, Irwing, Cammock & Lynn, 2001) wird in Chamorro-Premuzic et al. (2006a) keine Zuordnung zu einem Fähigkeitskonstrukt vorgenommen. Nach der CHC-Theorie, „knowledge of the culture“ erfolgte eine Zuordnung zu Gc (vgl. McGrew, 2009).

Durch die Berücksichtigung weiterer Studien und einer nachträglichen Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) konnten insgesamt mehr Datensätze kodiert werden ( $k = 764$ ) als in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien berücksichtigt wurde ( $k = 521$ ; Tabelle 6).

### 4.3 Fixed Effects Model vs. Random Effects Model

Das „Fixed Effects Model“ (FE-Modell) und „Random Effects Model“ (RE-Modell) sind bezüglich ihrer Eignung und Genauigkeit zur Bestimmung metaanalytischer Kennwerte kontrovers diskutiert worden (bspw. Field, 2001; Hedges & Vevea, 1998; Overton, 1998; Hunter & Schmidt, 2000). Unabhängig ihrer inhaltlichen Vor- und Nachteile konnte eine häufigere Verwendung des FE-Modells in den 90er Jahren im Vergleich zu dem RE-Modell beobachtet werden (Cooper, 1997; National Research Council, 1992). Als Grund für die höhere Frequenz der Nutzung des FE-Modells schreibt Cooper (1997, S. 179): „In practise, most meta-analysts opt for the fixed effects assumption because it is analytically easier to manage“ oder „...many users of meta-analysis prefer FE models because of the conceptual and computational simplicity“ (National Research Council, 1992, S. 52). Mittlerweile existieren jedoch Software-Pakete, bspw. von Schmidt und Le (2004), die die Durchführung einer Metaanalyse nach dem RE-Modell erheblich erleichtern.

Ein wesentlicher Unterschied zwischen dem FE-Modell und dem RE-Modell ist, dass das FE-Modell keine Schätzung der Varianz der wahren Effektgröße ( $S^2_{\rho}$ ) zulässt, weil eine Homogenität der Populationseffektgrößen a priori mit  $S^2_{\rho} = 0$  vorausgesetzt wird. Hingegen lässt das RE-Modell die Möglichkeit einer Variation der Populationsparameter in den verschiedenen Studien zu. Damit ist das RE-Modell das allgemeinere Modell, was bei der Schätzung der mittleren Effektstärke den Alpha-Fehler adäquat auf einem Signifikanzniveau von bspw. 5% halten kann, unabhängig davon, ob die Populationsparameter in allen Studien gleich sind oder variieren. Das FE-Modell ist hingegen eine spezielle und vereinfachte Version des RE-Modells mit  $S^2_{\rho} = 0$ . Denn wenn ein RE-Modell verwendet wird, bei dem  $S^2_{\rho} = 0$  ist, ist es mathematisch ein FE-Modell (Hunter & Schmidt, 2004). Wird allerdings ein FE-Modell mit  $S^2_{\rho} \neq 0$  verwendet, kann dies letztlich zu

Ungenauigkeiten der metaanalytischen Kennwerte führen: „...when population parameters vary across studies, the FE formulas have Type I error rates that are higher than the nominal values – often much higher“ (Hunter & Schmidt, 2004, S. 394). Überdies konnte gezeigt werden, dass das Konfidenzintervall des FE-Modells, welches auf der Grundlage des Standardfehlers berechnet wird, zu eng bemessen ist (Hunter & Schmidt, 2000, 2004). In dem Fall kann ein nominelles Konfidenzintervall von 95% einem tatsächlich substanziell geringeren Konfidenzintervall von bspw. 60% entsprechen. Die Vorgabe einer Präzision, die tatsächlich nicht vorhanden ist, würde hier zu einer Überschätzung der wahren Effektgröße führen.

Wenn bei der Verwendung des FE-Modells statistische Kennwerte wie bspw. das Konfidenzintervall oder der Populationszusammenhang berechnet werden, wird vorausgesetzt, dass die Studienpopulationsparameter aus derselben Grundgesamtheit stammen bzw. homogen sind. Zur Überprüfung der Homogenität wird der  $\chi^2$ -Test verwendet. Wenn der  $\chi^2$ -Test nicht signifikant ist, wird eine Homogenität der Studienpopulationsparameter angenommen. Allerdings weisen Hedges und Olkin (1985), Hunter und Schmidt (2004) und National Research Council (1992) darauf hin, dass der  $\chi^2$ -Test die Variation der Studienpopulationsparameter unzureichend anzeigt, insbesondere wenn die Anzahl der inkludierten Studien kleiner als 30 ist. Dementsprechend würde häufiger der Beta-Fehler auftreten. Eine Homogenität der Studienpopulationsparameter würde fälschlicherweise angenommen werden, obwohl tatsächliche Variationen bestehen. Wenn eine Homogenität fälschlicherweise angenommen wird, führt das zu einer Überschätzung des Populationszusammenhangs (siehe oben).

Trotz zunehmender Konvergenzen der interessierenden Konstrukte im Intelligenz- und Persönlichkeitsbereich (siehe Theorieteil), ist davon auszugehen, dass aufgrund der Zuordnung unterschiedlicher diagnostischer Verfahren zu den Konstrukten, Variationen in den Studienpopulationsparametern höchst wahrscheinlich sind. Dementsprechend wurde ein RE-Modell zu Effektgrößenbestimmung verwendet. Selbst wenn  $S^2_{\rho} = 0$  wäre, wäre es unproblematisch ein RE-Modell zu verwenden, da das RE-Modell in dem Fall mathematisch ein FE-Modell sein würde (siehe oben; Hunter & Schmidt, 2004). Für ausführlichere Darstellungen der unterschiedlichen Modelle zur Effektgrößenbestimmung in Metaanalysen sei bspw. auf Field (2001), Hedges und Vevea (1998), Hunter und Schmidt (2000, 2004), National Research Council (1992) sowie Overton (1998) verwiesen.

#### 4.4 Artefaktkorrekturen

Wenn alle berücksichtigten Studien in Metaanalysen die gleiche Stichprobengröße, die gleiche Reliabilität, die gleiche Validität und die gleiche Standardabweichung hätten, dann müssten die interessierenden Kennwerte der inkludierten Studien einfach gemittelt werden, um ein genaues Ergebnis über die Studien hinweg zu erhalten. Allerdings unterscheiden sich Studien in vieler Hinsicht und unterliegen einer Vielzahl von Artefakten in unterschiedlicher Ausprägung. Artefakte werden dabei als Mängel in Studien betrachtet, die sich auf die Genauigkeit der Ergebnisse in Metaanalysen nachteilig auswirken können: „We refer to study imperfections as 'artifacts' to remind ourselves that errors in study results produced by study imperfections are artifactual or man-made errors and not properties of nature“ (Hunter & Schmidt, 2004, S. 33).

Hunter und Schmidt (2004) berichten insgesamt 11 unterschiedliche Artefakte, die die Höhe der Korrelation der einzelnen Studien verändern können. Dabei wird angenommen, dass das Artefakt mit dem stärksten Verzerrungspotenzial der Stichprobenfehler ist (Hunter & Schmidt, 2004). Die ausschließlich um den Stichprobenfehler bereinigten Kennwerte werden in dem Software-Paket von Schmidt und Le (2004) als „Bare-Bones“-Analysen ausgegeben. Für eine ausführliche Darstellung der Korrektur des Stichprobenfehlers, einschließlich der zugrundeliegenden Formeln, sei bspw. auf Hunter und Schmidt (2004, S. 81-88) verwiesen.

Das bei metaanalytischen Untersuchungen Studien mit größeren Stichproben höher gewichtet werden als Studien mit kleineren Stichproben, kann zur genauen Bestimmung von Effektgrößen in der Population von Vorteil sein, da sich der Stichprobenfehler mit der Zunahme der Stichprobengröße verringert. Allerdings kann die Gewichtung von Effektgrößen anhand der Stichprobengrößen auch problematisch sein (Hunter & Schmidt, 2004). Nämlich dann, wenn eine einzelne oder wenige Stichproben mit sehr großen Stichprobengrößen mit der Höhe des Populationszusammenhangs korrelieren (Hunter & Schmidt, 2004). Das folgende Beispiel soll diese Problematik verdeutlichen. Wenn von 13 inkludierten Stichproben eine einzelne Studie eine Stichprobengröße von  $n = 15,000$  hat und die restlichen 12 Studien Stichproben von  $n < 500$  haben, würde die Stichprobe mit  $n = 15,000$  mehr als eine 30fache Gewichtung im Vergleich zu den restlichen 12 Studien erhalten (Hunter & Schmidt, 2004). Die metaanalytischen Befunde wären somit nahezu von der Effektstärke einer einzelnen Studie abhängig. Da allerdings nicht ausgeschlossen werden kann, dass auch eine Studie mit einer sehr großen Stichprobe Artefakten unterliegt, schlagen Hunter und Schmidt (2004, S. 82) vor: „...we recommend two analyses: a first analysis with the large-study included and a second analysis with the large-sample study left out“.



In Tabelle 7 sind die inkludierten Studien der vorliegenden Arbeit mit Stichproben  $n > 1,000$  aufgeführt. Das kleinste  $n$  in Tabelle 7 beträgt 1,821 (Furnham, Dissou, Sloan, Chamorro-Premuzic, 2007). Der Abstand zu der Studie mit der nächstkleineren Stichprobengröße beträgt  $n = 921$  (Moutafi, Furnham & Crump, 2003).

**Tabelle 7**

Inkludierte Studien aus der in der vorliegenden Arbeit durchgeführten Metaanalyse mit  $n > 1,000$

| <b>publizierte Studie</b>                           | <b>Stichprobengröße</b> |
|---|-------------------------|
| De Fruyt, Garcia, Rolland und Jung (2006)           | 20,183                  |
| Furnham, Dissou, Sloan und Chamorro-Premuzic (2007) | 1,821                   |
| Furnham, Jensen und Crump (2008)                    | 2,982                   |
| Furnham, Moutafi und Paltiel (2005)                 | 4,547                   |
| Moutafi, Furnham und Crump (2006)                   | 2,658                   |
| Noftle und Robins (2007)                            | 10,497                  |
| Reeve, Meyer und Bonaccio (2006)                    | 71,887                  |
| Schurr, Ruble und Henriksen (1988)                  | 1,902                   |

Wie bereits oben berichtet, beträgt die gesamte Stichprobengröße aller inkludierten Stichproben in der vorliegenden Arbeit  $N = 154,721$  mit 161 verschiedenen Datensätzen. Die durchschnittliche Stichprobengröße beträgt  $\bar{n} = 961$ . Zum Vergleich: Werden die Studien von De Fruyt et al. (2006;  $n = 20,183$ ) und Reeve et al. (2006;  $n = 71,887$ ) exkludiert, verringert sich die Gesamtstichprobe auf  $N = 62,651$ . Die daraus resultierende durchschnittliche Stichprobengröße würde sich um 59% auf  $\bar{n} = 394$  verringern. Um den überproportionalen Einfluss von zwei Effektgrößen zu kontrollieren, wurde in der vorliegenden Arbeit dem Vorschlag von Hunter und Schmidt (2004) gefolgt und zwei Analysen, mit und ohne Inklusion der Studien von De Fruyt et al. (2006) und Reeve et al. (2006), durchgeführt.

Neben der Korrektur des Stichprobenfehlers und der besonderen Berücksichtigung der Studien mit sehr großen Stichprobengrößen gibt es weitere Artefakte die einen systematischen Einfluss auf die Korrelationen haben können. Ein weiteres Artefakt ist die Unreliabilität der Messungen. Reliabilitäten sind von der Stichprobencharakteristik abhängig. Deshalb sind die Reliabilitäten am präzisesten, die anhand der Stichprobe ermittelt wurden, die auch in der jeweiligen Studie untersucht wurde (Rouse, 2007). Häufig werden allerdings keine Reliabilitäten anhand der Stichproben berichtet. In metaanalytischen Untersuchungen zur Reliabilität („Reliability Generalization“) wird berichtet, dass zwischen 3.2% und 63.6% nutzbare Reliabilitätskoeffizienten aus den inkludierten Einzelstudien angegeben sind (Caruso, 2000; Leue & Lange, 2011; Vacha-

Haase, Henson & Caruso, 2002; Youngstrom & Green, 2003). Dementsprechend häufig sind bei der Durchführung einer Metanalyse keine präzisen Reliabilitätsangaben verfügbar.

Wie in anderen Metaanalysen (bspw. Leue & Beauducel, 2008; Leue & Lange, 2011) wurde in der vorliegenden Arbeit Cronbachs Alpha als Maß für die Reliabilität berücksichtigt. Wenn Cronbachs Alpha anhand der Studien nicht berichtet waren<sup>1</sup>, wurden die berechneten Cronbachs Alpha aus einer metaanalytischen reliabilitätsgeneralisierenden Studie (RG-Studie) für NEO-Inventare (NEO-PI, NEO-PI-R, NEO-FFI, NEO-Kurzinventare; Caruso, 2000) entnommen. Der Vorteil dieser und anderer RG-Studien sind berechnete Populationsreliabilitätskoeffizienten, die geringeren Artefakten einzelner Studien unterliegen und so generellere Aussagen in der Population erlauben sollen (Leue & Lange, 2011; Rodriguez & Maeda, 2006). Für kognitive Leistungstests gibt es bis zum jetzigen Zeitpunkt keine verfügbare RG-Studie. Deshalb wurden die Cronbachs Alpha, wenn diese nicht anhand der Stichprobe berichtet wurden<sup>2</sup>, substituiert. Die Substitution erfolgte durch den gebildeten Mittelwert von .85, der anhand der berichteten Cronbachs Alpha der kognitiven Leistungstests aus den inkludierten Studien gebildet wurde.

Ein weiteres Artefakt sind Varianzunterschätzungen bzw. -überschätzungen („range restriction“; „range enhancement“) von Effektgrößen. Als Maß für die beobachteten Varianzunterschiede ( $u_x$ ) wird der Quotient aus der Standardabweichung der Stichprobe ( $SD_{SP}$ ) und der Population ( $SD_{Pop}$ ) gebildet (Hunter & Schmidt, 2004; S. 214):

$$u_x = SD_{SP} / SD_{Pop} \quad (2)$$

Oftmals werden Varianzunterschätzungen oder –überschätzungen im Prädiktor korrigiert (Hunter & Schmidt, 2004; Leue & Beauducel, 2008; Leue & Lange, 2011). In der vorliegenden Arbeit wurde ebenfalls der Prädiktor, mittels der beobachteten Varianzunterschiede der Stichproben-Intelligenzwerte, korrigiert. Die Standardabweichung der Stichproben-Intelligenzwerte wurde den inkludierten Studien entnommen, soweit diese berichtet waren. Die  $SD_{Pop}$  wurden aus den entsprechenden Testmanualen entnommen, da angenommen werden kann, dass die Normstichprobe des Testmanuals keinen bzw. geringen Varianzunterschätzungen oder –überschätzungen unterliegt (bspw. Hunter & Schmidt, 2004; Leue & Lange, 2011).

---

<sup>1</sup> Bspw. Allik und Realo (1997), Arteche, Chamorro-Premuzic, Furnham und Crump (2008), DeYoung, Peterson und Higgins (2005), Furnham, Zhang und Chamorro-Premuzic (2006), LePine, Colquitt und Erez (2000), Parkinson und Taggar (2006), Taggar, Hackett und Saha (1999), Ziegler, Schmukle, Egloff und Bühner (2010).

<sup>2</sup> Bspw. Austin et al. Gibson (1997), Beauducel et al. (2007), Christopher et al. (2010), Farsides und Woodfield (2003), Gow, Whiteman, Pattie und Deary (2005), Miller und Tal (2007), Reddock, Biderman und Nguyen (2011), Zimprich, Allemann und Dellenbach (2009).

Konnte die Varianzunterschätzung bzw. –überschätzung im Prädiktor für eine Studie nicht ermittelt werden, wurde sie anhand der gemittelten  $u_x$ -Werte der anderen Studien ( $k = 104$ ) mit .96 geschätzt. Gründe für eine Schätzung waren:

- In 35 Studien (22.72%) sind keine Standardabweichungen der Intelligenzwerte berichtet<sup>3</sup>.
- Ein nicht standardisierter Mittelwert mit dazugehöriger Standardabweichung wurde aus mehreren unterschiedlichen kognitiven Leistungstests gebildet bzw. berichtet (bspw. Maurer, Lippstreu & Judge, 2008; Silvia & Sanders, 2010).
- Testmanuale waren nicht verfügbar, bspw. ein estländischer Intelligenztest aus Allik & Realo (1997).
- Ein standardisierter Mittelwert mit dazugehöriger Standardabweichung wird berichtet, bspw. von  $M = 106.86$ ,  $SD = 6.94$  (Ku & Ho, 2010). Allerdings kann nicht bestimmt werden, ob es sich um einen IQ- oder SW-Wert handelt, da in dem Fall das Testmanual auf chinesisches war.

#### 4.5 Generalisierbarkeit von metaanalytischen Befunden

Nach Hunter und Schmidt (2004) werden Artefaktkorrekturen auch durchgeführt, um der Frage nachzugehen, ob der korrigierte Effekt über verschiedene Einflüsse hinweg generalisierbar ist. Da jedoch nicht alle Artefakte korrigiert werden können, empfehlen Hunter und Schmidt (2004, S. 146) die „75%-Regel“: „We were arguing that in any given meta-analysis it is probably the case that the unknown and uncorrected artifacts account for 25% of the variance“. Beträgt die artefaktkorrigierte Varianz des korrigierten Effektes weniger als 75%, ist nicht auszuschließen, dass die Höhe des Zusammenhanges ein Effekt mehrerer Subpopulationen ist. In dem Fall ist davon auszugehen, dass zusätzliche Moderatoren den Effekt erklären.

Wenn die erklärte Varianz des korrigierten Effektes in den metaanalytischen Berechnungen der vorliegenden Arbeit kleiner als 75% war, wurden folgende Moderatoren untersucht<sup>4</sup>:

---

<sup>3</sup> Bates und Shieles (2003), Bratko, Chamorro-Premuzic und Saks (2006), Chamorro-Premuzic und Furnham (2008), Chamorro-Premuzic, Furnham und Petrides (2006), Chamorro-Premuzic et al. (2005), Djapo, Kolenovic-Djapo, Djokic und Fako (2011), Francis (1997), Furnham (2005), Furnham und Chamorro-Premuzic (2004b), Furnham, Jensen und Crump (2008), Furnham, Moutafi und Chamorro-Premuzic (2005), Furnham, Moutafi und Paltiel (2005), Furnham, Taylor und Chamorro-Premuzic (2008), Furnham, Zhang und Chamorro-Premuzic (2006), Gaygisiz (2009), Gentry, Polzine und Wakefield Jr (1985), George, Stickle, Rachid und Wopnford (2007), Gow et al. (2005), Gregory, Nettelbeck und Wilson (2010), Hill, Turiano, Hurd, Mroczek und Roberts (2011), Judge, Higgins, Thoresen und Barrick (1999), Kanfer, Ackerman und Heggestad (1996), Kaufman et al. (2010), Lounsbury, Sundstrom, Loveland und Gibson (2003), McCrae (1992), McCrae und Costa (1985), Miller und Tal (2007), Moutafi et al. (2003), Moutafi, Furnham und Paltiel (2004), Pincombe, Luciano, Martin und Wright (2007), Schurr, Ruble und Henriksen (1988), Sharp, Reynolds, Peterson und Gatz (2010), Simon, Fontana und Williams (1986), Williams, Nathanson und Paulhus (2010), Zimprich, Allemand und Dellenbach (2009).

- NEO- bzw. Costa und McCrae-Inventare: NEO-PI-(R) vs. NEO-FFI,
- NEO- bzw. Costa und McCrae-Inventare vs. andere FFM Inventare, bspw. International Personality Item Pool (Goldberg, 1992, 1999), Personal Characteristics Inventory (Barrick & Mount, 1995), Vijf Persoonlijkheids-Factoren Test (Elshout & Akkerman, 1975), Perspectives Inventory (Abraham & Morrison, 2002),
- FFM-Inventare vs. andere Inventare, die Ne und E erfassen, bspw. Eysenck Personality Questionnaire (Eysenck & Eysenck, 1975), Eysenck Personality Profiler (Eysenck & Wilson, 1991),
- Wonderlic Personnel Test (WPT; Wonderlic, 1992) vs. andere kognitive Leistungstests, die die Allgemeine Intelligenz erfassen, bspw. Wechsler Adult Intelligence Scale (Wechsler, 1981),
- Baddeley Reasoning Test (BRT; Baddeley, 1968) vs. Raven Tests (bspw. Raven, Raven & Court, 2000) vs. andere kognitive Leistungstests, die die fluide Intelligenz erfassen, bspw. Culture Fair Intelligence Test (Cattell, 1973),
- General Knowledge Test (GKT; Irwing, Cammock & Lynn, 2001) vs. andere kognitive Leistungstests, die Gc erfassen, bspw. Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1964).

Um die Generalisierbarkeit der metaanalytischen Befunde zu überprüfen, empfehlen Hunter und Schmidt (2004) neben der Berücksichtigung der „75%-Regel“ die Interpretation des Glaubwürdigkeitsintervalls („Credibility Interval“, CRI). Das CRI wird mit Hilfe der korrigierten Standardabweichung um den mittleren korrigierten Effekt gebildet. Ist das CRI sehr weit, kann nicht ausgeschlossen werden, dass die artefaktkorrigierte Effektgröße ein mittlerer Effekt mehrerer Subpopulationen ist (Whitener, 1990). Das Software-Paket von Schmidt und Le (2004), welches in der vorliegenden Arbeit zur Bestimmung metaanalytischer Kennwerte verwendet wurde, legt automatisch ein 80%iges CRI an.

Die CRIs und die „75%-Regel“ geben beide Antworten auf die Frage der Generalisierbarkeit metaanalytischer Befunde. Allerdings sind die Antworten die sie geben unter Umständen unterschiedlich. Berücksichtigend, dass das CRI liberaler ist, dementsprechend leichter eine homogene Population nahe legt, während die „75%-Regel“ vergleichsweise konservativer ist (Kramer, 2009), wurde in der vorliegenden Arbeit nach Kemery et al. (1987) eine Interpretation

---

<sup>4</sup> Moderatoranalysen wurden durchgeführt, wenn einem Moderator mindestens eine Stichprobenanzahl von  $k = 5$  zugrunde lag. Durch eine inkludierte Studieanzahl von  $k \geq 5$  sollte die Robustheit der Befunde gewährleistet sein.

der Generalisierbarkeit der Befunde anhand der Kombination des CRI und der „75%-Regel“ durchgeführt:

- Wenn die Null nicht im CRI liegt und mehr als 75% der Varianz der korrigierten Effektstärke erklärt wird, kann der artefaktkorrigierte Populationszusammenhang generalisiert werden.
- Liegt hingegen die Null im CRI und werden weniger als 75% der Varianz erklärt, liegt mit hoher Wahrscheinlichkeit ein mittlerer Effekt heterogener Subpopulationen vor, bei der die Höhe und Richtung nicht verallgemeinert werden sollten.
- Liegt die Null nicht im CRI und es werden weniger als 75% der Varianz erklärt, kann die Richtung des artefaktkorrigierten Populationszusammenhangs verallgemeinert werden. Allerdings sollte in dem Fall nicht ausgeschlossen werden, dass die Höhe des artefaktkorrigierten Populationszusammenhangs von Moderatoren abhängt (Kemery et al., 1987).

## 4.6 Spezifikation verwendeter Methoden zur Identifikation des Publikationsbias

### 4.6.1 Identifikation und Adjustierung des Publikationsbias nach der „trim and fill“-Methode

Die „trim and fill“-Methode nach Duval und Tweedie (2000a, 2000b) ist ein Verfahren mit dessen Hilfe ein Publikationsbias aufgrund einer asymmetrischen Verteilung publizierter Effektgrößen um die mittlere Effektgröße identifiziert werden soll. Besteht die Annahme einer asymmetrischen Verteilung, wird durch einen „trim“-Schritt (Abgleichschritt) die Effektstärke exkludiert, die am weitesten über den mittleren Effekt liegt. Der Abgleichschritt wird so lange wiederholt, bis eine Symmetrie der Effektgrößen hergestellt ist. Die daraus resultierende mittlere Effektgröße soll die Korrelation ohne Publikationsbias schätzen, führt aber gleichzeitig zu einer Unterschätzung der Varianz der mittleren Effektgröße (Duval & Tweedie, 2000a, 2000b). Deshalb sollen in einem darauffolgenden „fill“-Schritt (Auffüllschritt) die zuvor exkludierten Effektgrößen zusammen mit den zusätzlich konvergierenden symmetrischen Werten inkludiert werden. Die mittlere Effektgröße bleibt dabei unverändert.

In Duval (2005) sowie Duval und Tweedie (2000a, 2000b) werden verschiedene Schätzer, nämlich  $R_0$ ,  $L_0$  und  $Q_0$  berichtet, mit deren Hilfe die Anzahl fehlender Studien bzw. die Höhe eines Bias angezeigt werden kann. Dabei sollen sich insbesondere  $R_0$  und  $L_0$  als gute Schätzer zur Bestimmung fehlender Studien bewährt haben (Duval, 2005). Aufgrund möglicher

Ungenauigkeiten des Schätzers  $Q_0$  wird dieser in der vorliegenden Arbeit nicht weiter berücksichtigt (vgl. Duval & Tweedie, 2000a, 2000b). Die Formeln zur Berechnung der Schätzer  $R_0$  und  $L_0$  sind im Anhang aufgeführt (siehe S. 254).

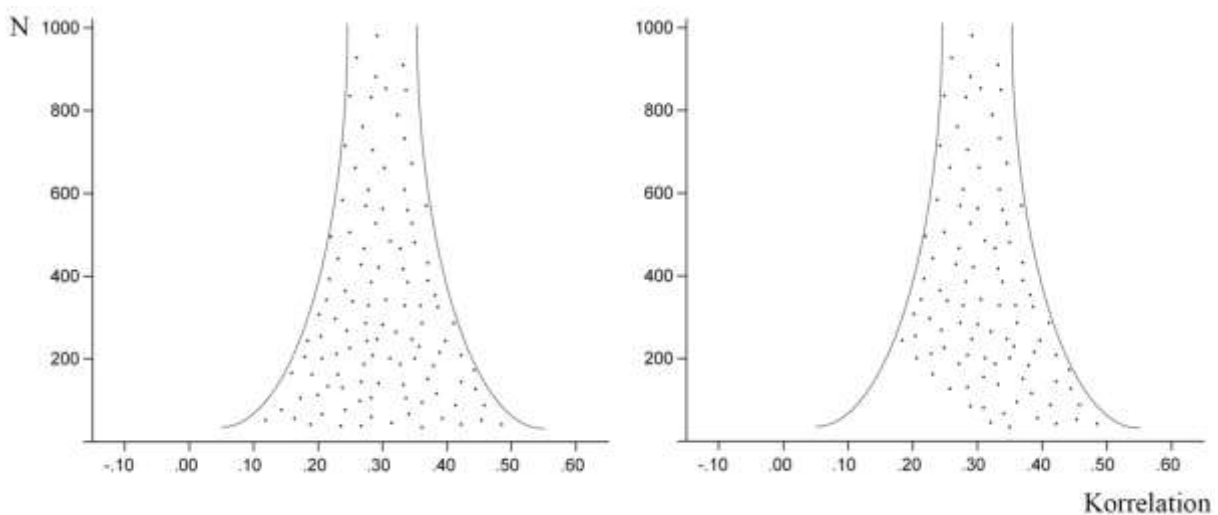
Ist  $R_0$  und/oder  $L_0 \geq 1$  soll auf die nächst größere Zahl aufgerundet werden. Die Schätzer werden nach der Aufrundung als  $R_0^+$  und  $L_0^+$  bezeichnet und geben an, wie viele Effektgrößen zu exkludieren sind. In Abhängigkeit der Anzahl zu exkludierender Effektgrößen werden die jeweils höchsten positiven Effektgrößen exkludiert (Duval, 2005). Danach wird der Mittelwert der Effektgrößen neu bestimmt ( $\theta_2$ ) und von  $r_i$  subtrahiert. Die Ränge werden wieder gebildet und  $R_0^+$  und  $L_0^+$  nach der oben beschriebenen Prozedur neu bestimmt. Dieser Vorgang wird wiederholt ( $\theta_j$ ), bis sich die Ränge bzw.  $L_0^+$  und  $R_0^+$  nicht mehr verändern (Formel 3). Anschließend wird die Anzahl von  $R_0^+$  oder/und  $L_0^+$  mit den konvergierenden symmetrischen Werten aufgefüllt („trim“-Schritt). Die Auffüllung mit den konvergierenden symmetrischen Werten ist für die mittlere Effektstärke nicht relevant, jedoch kann so die unterschätzte Varianz korrigiert werden (Duval & Tweedie, 2000a, 2000b).

$$r_j = \left\{ \left\{ \begin{array}{l} r_i - \theta_j \text{ if } r_i < 0 \\ \theta_j - r_i \text{ if } r_i > 0 \end{array} \right\}; (i = 1, 2, \dots, k_j, j = 1, 2, \dots, k_j) \right\} \quad (3)$$

Aufgrund der Annahme, dass bei einem Publikationsbias tendenziell mehr Effektgrößen mit einer höheren positiven Korrelation publiziert sind als mit kleineren positiven oder negativen Effektgrößen, werden nach der „trim and fill“-Methode entsprechend die extremsten positiven Effektgrößen exkludiert. Allerdings kann nicht ausgeschlossen werden, dass es auch einen Publikationsbias in die andere Richtung geben könnte. In dem Fall würden tendenziell mehr Effektgrößen mit höheren negativen Effektgrößen publiziert sein. Dementsprechend sollten die extremsten negativen Effektgrößen exkludiert werden. Jedenfalls sollte, bevor mit der Exklusion einzelner Effektgrößen begonnen wird, eine sorgfältige Beurteilung mit den gesamten publizierten Effektgrößen durchgeführt werden. Kann bspw. durch eine visuelle Prüfung des Streudiagramms, eine Tendenz der Fluktuation positiver oder kleiner negativer Effektgrößen angenommen werden, sind entsprechend die höchsten negativen Effektgrößen zu exkludieren und im Auffüllschritt die konvergierenden symmetrischen Werte, zu inkludieren (die Formeln müssen entsprechend angepasst werden).

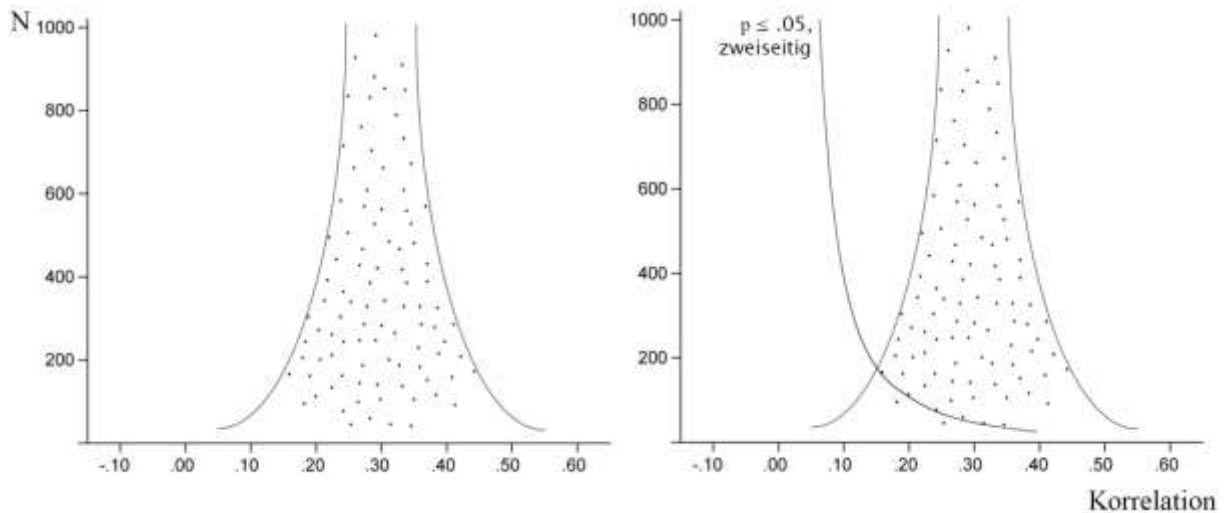
#### 4.6.2 Identifikation möglicher Publikationsverzerrungen mittels Signifikanzfunktion im Streudiagramm

Neben der oben vorgestellten Möglichkeit zur Identifikation eines Publikationsbias gibt es auch grafische Methoden, um asymmetrische Verteilungen festzustellen. Eine grafische Methode haben Light und Pillemer (1984) vorgeschlagen, in der mit Hilfe eines umgekehrten Trichters im Streudiagramm („Funnel-Plot“) eine asymmetrische Verteilung aufgrund der Anordnung der Effektgrößen festgestellt werden kann (Abbildung 8). „Funnel-Plots“ berücksichtigen die Annahme, dass Studien mit kleinerer Stichprobengröße eine größere Streuung der Effektgrößen haben, jedoch mit Zunahme der Stichprobengröße die Streuung der Effektgrößen abnimmt. Mit Hilfe eines Streudiagramms und eingezeichnetem umgekehrten Trichter kann durch visuelle Prüfung eine asymmetrische Verteilung identifiziert werden (Abbildung 8).



**Abbildung 8.** Beispiel eines (fiktiven) symmetrischen „Funnel-Plot“ (links) und eines (fiktiven) asymmetrischen „Funnel-Plot“ (rechts).

Besteht wie in Abbildung 8 (rechts) eine eindeutig asymmetrische Verteilung der Effektgrößen, kann im linken Bereich, innerhalb des eingezeichneten umgekehrten Trichters bei Stichprobengrößen mit  $n < 200$ , ein Publikationsbias nicht ausgeschlossen werden. Es kann jedoch auch vorkommen, dass sich die Effektgrößen symmetrisch verteilen und dennoch die Annahme eines Publikationsbias nicht ausgeschlossen werden sollte (Abbildung 9, links). In dem Fall wäre der Auflösungsgrad eines „Funnel-Plot“ möglicherweise nicht ausreichend, um weitere präzise Informationen über ein Publikationsbias zu erhalten.



**Abbildung 9.** Beispiel eines (fiktiven) „Funnel-Plot“ ohne eindeutigen Hinweis auf eine asymmetrische Verteilung (links) und mit zusätzlich eingezeichneter Signifikanzfunktion ( $p \leq .05$ , zweiseitig; rechts).

In Abbildung 9 sind bei Stichproben mit Stichprobengrößen  $n < 200$  innerhalb des eingezeichneten, umgekehrten Trichters keine Effektgrößen im linken und rechten Bereich abgebildet bzw. publiziert. Dementsprechend wäre ein Publikationsbias im linken und/oder rechten Bereich, innerhalb des eingezeichneten umgekehrten Trichters bei  $n < 200$  anzunehmen (Abbildung 9, links). Dennoch kann eine symmetrische, wenn auch in Bezug zum eingezeichneten Trichter, fehlende Werte aufweisende Verteilung anhand einer visuellen Prüfung angenommen werden. In dem dargestellten Fall würden Methoden, die ein Publikationsbias aufgrund einer asymmetrischen Verteilung der Effektgrößen annehmen (bspw. Duval & Tweedie, 2000a, 2000b), kein Publikationsbias anzeigen. Um die Möglichkeit Informationen bezüglich spezifischer Veröffentlichungstendenzen mit Hilfe einer visuellen Prüfung im Streudiagramm zu erhöhen, könnte es sich anbieten, eine Funktion einzuzichnen, die die Grenze für signifikante Korrelationen auf dem 5%-Niveau (zweiseitig) markiert (Abbildung 9, rechts). Anhand der Funktion wird deutlich, dass im gesamten, nicht signifikanten Bereich ( $p > .05$ , zweiseitig) nur zwei Effektgrößen publiziert sind. Hingegen sind in dem fiktiven Beispiel sechs Effektgrößen auf bzw. nahe der eingezeichneten Signifikanzfunktion im signifikanten Bereich ( $p \leq .05$ , zweiseitig) publiziert. Neben dieser ungleichen Verteilung der abgebildeten Effektgrößen nahe der Signifikanzgrenze ist auffällig, dass sich die sechs signifikanten Effektgrößen entlang der dargestellten Signifikanzgrenze verteilen. Aufgrund dieser zusätzlichen Information, könnte ein Publikationsbias im linken Bereich bei  $n < 200$  angenommen werden (Abbildung 9, rechts). Eine derartige Verteilung von Effektgrößen könnte bspw. eintreten, wenn Stichprobengrößen systematisch erhöht werden, um signifikante Korrelationen zu erhalten.



In Abhängigkeit von der Fragestellung der verwendeten Effektgrößen ( $r$ ,  $d$ , etc.) und der Datenlage könnten unterschiedliche Signifikanzfunktionen in ein Streudiagramm eingezeichnet werden. Überdies wären auch unterschiedliche eingezeichnete Signifikanzfunktionen für verschiedene Signifikanzniveaus möglich.

Das Einzeichnen eines umgekehrten Trichters in ein Streudiagramm nach Light und Pillemer (1984) ist eine grafische Möglichkeit, um ein Publikationsbias mittels einer asymmetrischen Verteilung publizierter Effektgrößen zu schätzen. Unter Umständen kann allerdings der Auflösungsgrad dieser Methode als unzureichend bewertet werden (siehe oben). Die Berücksichtigung einer Signifikanzfunktion im Streudiagramm bietet zusätzlich die Möglichkeit, Veröffentlichungstendenzen entlang der Signifikanzgrenze visuell zu prüfen. Dennoch könnte sich beim Auftreten komplexer Verteilungen publizierter Effektgrößen die Methode des eingezeichneten umgekehrten Trichters und der eingezeichneten Signifikanzfunktion als unzureichend erweisen. Deshalb werden im Folgenden Überlegungen beschrieben, wie man mit Hilfe von möglichst exakten Kennwerten für die Verteilung der Effektgrößen entlang der Signifikanzfunktion, die Wahrscheinlichkeit erhöhen könnte, ein Publikationsbias zu identifizieren.

### **4.6.3 Index zum Verhältnis signifikanter und nichtsignifikanter Effektgrößen**

Das Verhältnis zwischen publizierten signifikanten und publizierten nichtsignifikanten Effektgrößen scheint ein wichtiger Aspekt bei der Untersuchung von Publikationsbias zu sein (Hedges, 1992; Hedges & Olkin, 1985; Iyengar & Greenhouse, 1988; Rosenthal, 1979). Um ein theoretisches Ideal der Zufallsverteilung von signifikanten und nichtsignifikanten Effektgrößen ohne ein Publikationsbias zu erhalten, bietet es sich an, eine Simulation durchzuführen. Über eine Simulation kann ein Referenzwert gebildet werden, der das Verhältnis zwischen signifikanten und nichtsignifikanten Effektgrößen – ohne Publikationsbias – anzeigt. Der aus der Simulation ermittelte Referenzwert kann mit dem Wert, der sich aus dem Verhältnis tatsächlich publizierter signifikanter und nichtsignifikanter Effektgrößen berechnet, verglichen werden. Besteht eine Differenz zwischen dem theoretischen Referenzwert und dem empirischen Wert, sollte ein Publikationsbias nicht ausgeschlossen werden. Der Gedanke, ein Publikationsbias mittels eines Vergleichs zwischen theoretischem Ideal und gegebenen empirischen Sachverhalt zu identifizieren, wird so quantifizierbar.

Aus der Verteilung signifikanter und nichtsignifikanter Effektgrößen kann ein Index berechnet werden, der das Verhältnis signifikanter und nichtsignifikanter Effektgrößen anzeigt. Soll der Index den prozentualen Anteil signifikanter Effektgrößen ( $p \leq .05$ , zweiseitig) anzeigen, wird die Anzahl signifikanter Effektgrößen dividiert durch die Summe der Anzahl signifikanter Effektgrößen (sE;  $p \leq .05$ , zweiseitig) und nichtsignifikanter Effektgrößen (nsE;  $p > .05$ , zweiseitig).

$$\text{Index} = \frac{\sum_i \text{sE}}{\sum_i (\text{sE} + \text{nsE})} \quad (4)$$

Der so berechnete Index hat einen Wertebereich von null bis eins, wobei null anzeigen würde, dass alle Effektgrößen nichtsignifikant sind ( $p > .05$ , zweiseitig) und ein Index von eins anzeigen würde, dass alle publizierten Studien signifikant wären ( $p \leq .05$ , zweiseitig). Dazu ein Beispiel: In Abbildung 9 (rechts) sind im signifikanten Bereich ( $p \leq .05$ , zweiseitig) 94 Effektgrößen und im nichtsignifikanten Bereich ( $p > .05$ , zweiseitig) zwei Effektgrößen abgebildet, was einem Index nach Formel 4 von .98 entsprechen würde. Demnach sind 98% signifikante ( $p \leq .05$  zweiseitig) und 2% nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) bei einer mittleren Effektgröße von  $\bar{r} = .30$  publiziert (Abbildung 9, rechts).

Es gibt verschiedene Parameter, die sich unabhängig eines Publikationsbias auf die Höhe des Index auswirken können. Ein Parameter, der sich auf das Verhältnis zwischen signifikanten und nichtsignifikanten Effektgrößen auswirken kann, ist die mittlere Effektgröße. Beträgt die mittlere Effektgröße bspw.  $\bar{r} = .00$ , können weniger signifikante als nicht signifikante Effektgrößen erwartet werden als bei einer mittleren Effektgröße von  $\bar{r} = .50$ . Dementsprechend ist bei der Interpretation des Index hinsichtlich eines Publikationsbias die mittlere Effektgröße zu berücksichtigen. Ein weiterer Parameter, der sich auf das Verhältnis zwischen signifikanten und nichtsignifikanten Effektgrößen – ohne Publikationsbias – auswirken kann, sind die Stichprobengrößen, die den Effektgrößen zugrunde liegen. Um den Einfluss der mittleren Effektgröße und der Stichprobengrößen publizierter Studien auf die Indexhöhe – ohne Publikationsbias – zu untersuchen, wurden Simulationen durchgeführt, die nachfolgend beschrieben werden.

### 4.6.4 Auswirkungen auf die Indexhöhe durch Parametervariationen

Simulationen bieten den Vorteil, dass Parameter wie die mittlere Effektgröße, die Anzahl der Effektgrößen und die Stichprobengrößen variiert werden können. Gleichzeitig kann ein

Publikationsbias ausgeschlossen werden. Mit der Methode der zufallsbedingten Erzeugung gewichteter Aggregationen (Beauducel, 2001a, 2001b; Knuth, 1981) kann in Abhängigkeit eines zuvor festgelegten Populationszusammenhangs eine hinreichend große Anzahl zufallsbedingter Effektgrößen erzeugt werden. Der daraus ermittelte Index nach Formel 4 sollte den prozentualen Wert signifikanter Effektgrößen ( $p \leq .05$ , zweiseitig) für einen Populationszusammenhang – ohne Publikationsbias – als ein theoretisches Ideal anzeigen.

Mit SPSS 16.0 wurden bivariate Produkt-Moment-Korrelationen für festgelegte Effektstärken simuliert. Dazu wurden jeweils zwei z-standardisierte Variablen, die in der Population von 200,000 Fällen eine bestimmte Korrelation hatten, generiert. Diese zwei korrelierten Variablen wurden anhand von drei normalverteilten, orthogonalen, z-standardisierten Zufallsvariablen generiert. Die erste Zufallsvariable  $c$  (= „common variance“) repräsentiert den Anteil gemeinsamer Varianz, die beiden anderen Zufallsvariablen repräsentieren den jeweils spezifischen Varianzanteil  $s_1$  bzw.  $s_2$ . Die Formeln für die Generierung der zwei Variablen mit einem Populationszusammenhang ( $\rho$ ) anhand einer gewichteten Aggregation (Beauducel, 2001a, 2001a; Knuth, 1981) sind

$$v_1 = \rho^{0.5}c + (1 - \rho)^{0.5}s_1, \quad (5)$$

$$v_2 = \rho^{0.5}c + (1 - \rho)^{0.5}s_2. \quad (6)$$

Das Gewicht für die Aggregation wurde hier als  $\rho$  bezeichnet, da das Gewicht genau dem Populationszusammenhang entspricht. Aus dieser Population zwei korrelierter Variablen mit 200,000 Fällen wurden jeweils 100 Zufallsstichproben mit Stichprobengrößen in 10er Schritten zwischen  $n = 30$  und  $n = 1,000$  Fällen gezogen. Die anhand der Zufallsstichproben ermittelten Korrelationen zwischen den beiden Variablen  $v_1$  und  $v_2$  variieren dann nur aufgrund des Stichprobenfehlers, wobei der Mittelwert der simulierten Korrelationen dem Populationszusammenhang sehr nahe kommt. Da die simulierten Korrelationen keinerlei Publikationsbias aufweisen, kann die Verteilung dieser Korrelationen als Referenzgröße bei der Ermittlung des Publikationsbias herangezogen werden.

Aufgrund 100 gezogener Zufallsstichproben mit Stichprobengrößen von jeweils  $n = 30$  bis  $n = 1,000$  in 10er Schritten wurden 9,800 Effektgrößen für einen einzelnen Populationszusammenhang berücksichtigt. Um die Veränderung des Index in Abhängigkeit des Populationszusammenhangs zu untersuchen, wurde der Populationszusammenhang insgesamt 12-mal in .05er Schritten zwischen  $\rho = .00$  und  $\rho = .55$  variiert. Nach Formel 4 wurde für jede einzelne der 12

## *Kapitel 4 – Methode*

Populationszusammenhänge der Index berechnet, der den prozentualen Anteil signifikanter Effektgrößen ( $p \leq .05$ , zweiseitig) im Vergleich zu den nicht signifikanten Effektgrößen ( $p > .05$ , zweiseitig) – ohne Publikationsbias – anzeigt (Tabelle 8, Simulation 1).

Um die Veränderung der Indexhöhe in Abhängigkeit der Stichprobengrößen zu untersuchen, wurden zwei weitere Simulationen durchgeführt. In der ersten der zwei zusätzlichen Simulationen (Tabelle 8, Simulation 2) wurde die maximale Stichprobengröße von  $n = 1,000$  auf  $n = 500$  verringert. Durch die Exklusion der Stichprobengrößen von  $n = 510$  bis  $n = 1,000$  verringerte sich die Anzahl der Stichproben von  $k = 9,600$  auf  $k = 4,800$ . Dadurch kann der Einfluss der Abnahme der maximalen Stichprobengröße bei gleichzeitiger Verringerung der Stichprobenanzahl ohne Annahme eines Publikationsbias auf den Index festgestellt werden (Tabelle 8, Simulation 2).

In einer weiteren Simulation (Simulation 3, Tabelle 8) wurde der Einfluss der Erhöhung der Mindeststichprobengröße im Vergleich zu Simulation 1 auf den Index untersucht. Dazu wurde die Mindeststichprobengröße von  $n = 30$  auf  $n = 70$  erhöht. Durch die Exklusion der Stichprobengrößen von  $n = 30$  bis  $n = 60$  verringert sich die Anzahl der Stichproben von  $k = 9,600$  auf  $k = 9,400$  (Tabelle 8, Simulation 3).

**Tabelle 8** Ermittelte Indizes in Abhängigkeit der Variation von  $\rho$ ,  $N$  und  $k$  – ohne Publikationsbias – mittels Simulation

|              | $\rho = .00$ |            |            | $\rho = .05$ |            |            | $\rho = .10$ |            |            | $\rho = .15$ |            |            | $\rho = .20$ |             |            | $\rho = .25$ |             |            |
|--------------|--------------|------------|------------|--------------|------------|------------|--------------|------------|------------|--------------|------------|------------|--------------|-------------|------------|--------------|-------------|------------|
|              | S1           | S2         | S3         | S1           | S2         | S3         | S1           | S2         | S3         | S1           | S2         | S3         | S1           | S2          | S3         | S1           | S2          | S3         |
| $\sum$ sE    | 221          | 110        | 217        | 1,700        | 511        | 1,686      | 5,021        | 1,457      | 4,995      | 7,592        | 2,824      | 7,545      | 8,665        | 3,765       | 8,582      | 9,147        | 4,247       | 9,019      |
| $\sum$ nsE   | 9,579        | 4,690      | 9,183      | 8,100        | 4,289      | 7,714      | 4,779        | 3,343      | 4,405      | 2,208        | 1,976      | 1,855      | 1,135        | 1,035       | 818        | 653          | 553         | 381        |
| <b>Index</b> | <b>.02</b>   | <b>.02</b> | <b>.02</b> | <b>.17</b>   | <b>.12</b> | <b>.18</b> | <b>.51</b>   | <b>.30</b> | <b>.53</b> | <b>.77</b>   | <b>.59</b> | <b>.80</b> | <b>.88</b>   | <b>.78</b>  | <b>.91</b> | <b>.93</b>   | <b>.88</b>  | <b>.96</b> |
|              | $\rho = .30$ |            |            | $\rho = .35$ |            |            | $\rho = .40$ |            |            | $\rho = .45$ |            |            | $\rho = .50$ |             |            | $\rho = .55$ |             |            |
|              | S1           | S2         | S3         | S1           | S2         | S3         | S1           | S2         | S3         | S1           | S2         | S3         | S1           | S2          | S3         | S1           | S2          | S3         |
| $\sum$ sE    | 9,398        | 4,496      | 9,189      | 9,537        | 4,637      | 9,270      | 9,619        | 4,719      | 9,290      | 9,669        | 4,769      | 9,299      | 9,684        | 4,784       | 9,300      | 9,694        | 4,794       | 9,300      |
| $\sum$ nsE   | 402          | 304        | 211        | 263          | 163        | 130        | 181          | 81         | 110        | 131          | 31         | 101        | 116          | 16          | 100        | 106          | 6           | 100        |
| <b>Index</b> | <b>.96</b>   | <b>.94</b> | <b>.98</b> | <b>.97</b>   | <b>.97</b> | <b>.99</b> | <b>.98</b>   | <b>.98</b> | <b>.99</b> | <b>.99</b>   | <b>.99</b> | <b>.99</b> | <b>.99</b>   | <b>1.00</b> | <b>.99</b> | <b>.99</b>   | <b>1.00</b> | <b>.99</b> |

Anmerkungen.  $\rho$  = Populationszusammenhang; S1 = Simulation 1, Stichprobengrößen von  $n = 30$  bis  $n = 1,000$ ; S2 = Simulation 2, Stichprobengrößen von  $n = 30$  bis  $n = 500$ ; S3 = Simulation 3, Stichprobengrößen von  $n = 70$  bis  $n = 1,000$ ;  $\sum$  sE = Summe der Anzahl signifikanter Effektgrößen ( $p \leq .05$ , zweiseitig);  $\sum$  nsE = Summe der Anzahl nichtsignifikanter Effektgrößen ( $p > .05$ , zweiseitig); Index =  $\sum$  sE /  $\sum$  (sE + nsE).

Mit einer fixierten Stichprobenobergrenze von  $n = 1,000$  in der Simulation 1 und 3 wird ab einem Populationszusammenhang von  $\rho = .10$  der Index von  $.50$  überschritten (Tabelle 8). Damit ein Publikationsbias auszuschließen wäre, müssten in dem Fall ungefähr 50% nicht signifikante Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Mit einer festgelegten Stichprobenobergrenze von  $n = 500$  und einem unveränderten Populationszusammenhang von  $\rho = .10$  beträgt der Index hingegen  $.30$  (Simulation 2, Tabelle 8). Durch die Exklusion der Stichprobengrößen von  $n = 510$  bis  $n = 1,000$  verringert sich der Index um 20%. Dementsprechend wären 70% nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) zu erwarten, damit ein Publikationsbias ausgeschlossen werden könnte.

Aufgrund dieser Befundlage ist davon auszugehen, dass sich die Höhe verschiedener zu inkludierender Stichprobengrößen auf den Index – ohne Publikationsbias und unverändertem Populationszusammenhang – auswirkt. Bei der Erzeugung bzw. Berücksichtigung von Stichproben mit großem  $n$ , nimmt die Wahrscheinlichkeit zu, vermehrt signifikante Effektgrößen zu erhalten. Allerdings wäre die daraus resultierende Erhöhung des Index kein Hinweis auf ein Publikationsbias, sondern schlicht eine Auswirkung der größeren Teststärke.

Darüber hinaus konnte ab einem Populationszusammenhang von  $\rho = .45$  ein Deckeneffekt des Index beobachtet werden (Tabelle 8). Der Index nähert sich hierbei dem Maximalwert von eins an, ohne dass ein Publikationsbias vorhanden wäre. Ähnlich verhält es sich bei einem Populationszusammenhang von  $\rho = .00$ . Die Indizes nähern sich dem Minimalwert von Null an, ohne dass ein Publikationsbias vorhanden wäre (Tabelle 8; Abbildung 10).

Aufgrund der durchgeführten Simulationen kann angenommen werden:

- Die Indizes unterscheiden sich in Abhängigkeit des Populationszusammenhangs<sup>5</sup>.
- Die Höhe inkludierter Stichprobengrößen wirkt sich bei Populationszusammenhängen zwischen  $|\rho| = .00$  und  $|\rho| = .45$  auf die Höhe des Index aus. Daher trägt bei einem  $|\rho| = .00$  oder  $|\rho| > .45$  die hier vorgestellte Methode nicht zur Identifikation eines Publikationsbias bei.

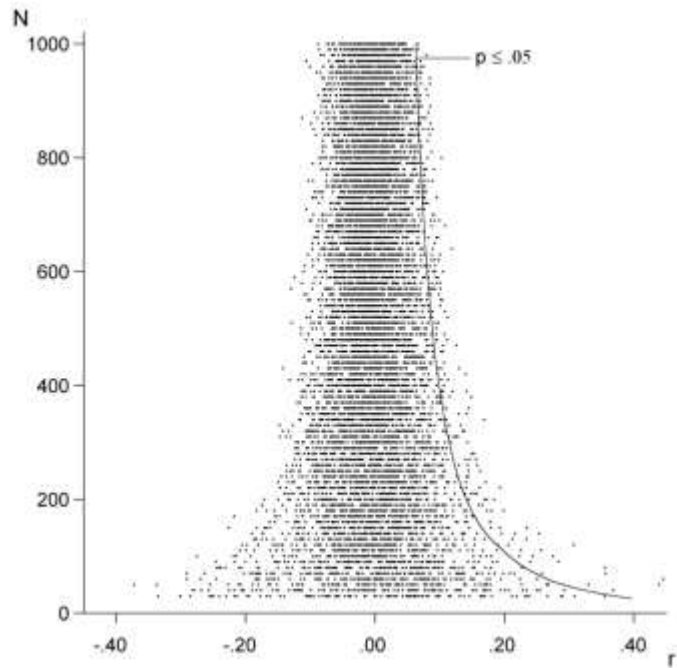
---

<sup>5</sup> Da sich der Populationszusammenhang auf das Verhältnis gezogener signifikanter ( $p \leq .05$ , zweiseitig) und nichtsignifikanter ( $p > .05$ , zweiseitig) Effektgrößen in der Simulation auswirkt, wurden die Studien mit sehr großen Stichprobengrößen (De Fruyt et al., 2006,  $n = 20,175$ ; Reeve et al., 2006,  $n = 71,887$ ) bei der Durchführung von Simulationen in der vorliegenden Arbeit nicht berücksichtigt. Dadurch soll vermieden werden, dass sich eine einzelne Studie aufgrund der Stichprobengröße möglicherweise überproportional auf den Populationszusammenhang und somit möglicherweise verzerrend auf das Verhältnis signifikanter ( $p \leq .05$ , zweiseitig) und nichtsignifikanter ( $p > .05$ , zweiseitig) Effektgrößen auswirkt.

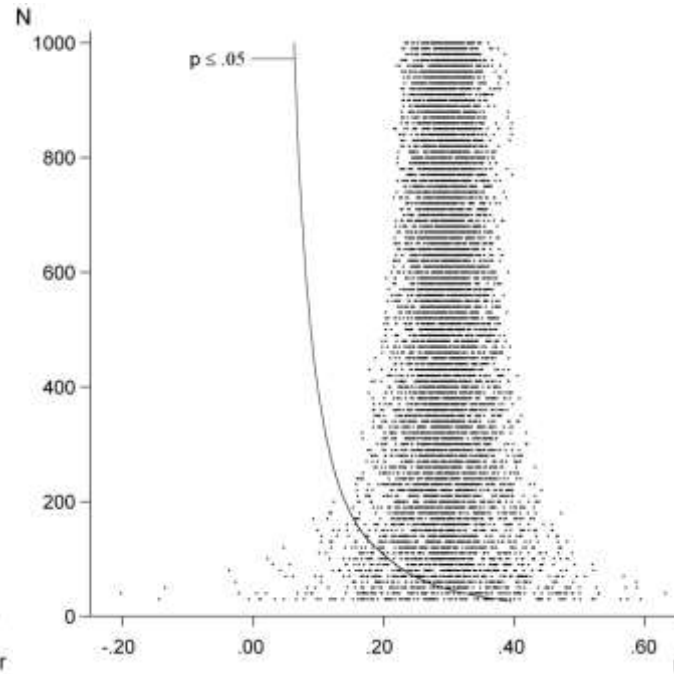
## *Kapitel 4 – Methode*

- Ein einzelner, allgemeingültiger Index ist unzureichend, um zu beurteilen, ab welcher Indexhöhe das Vorhandensein ein Publikationsbias sehr wahrscheinlich ist.

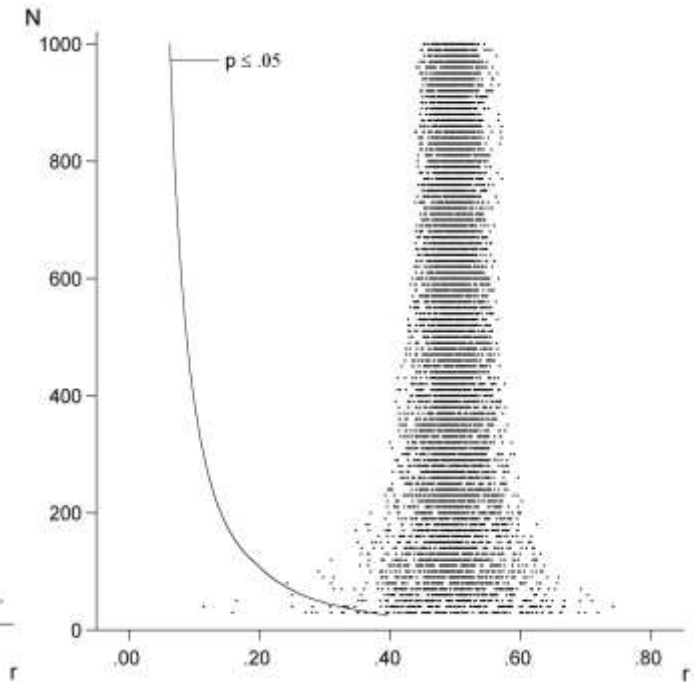
Um die Veränderung des Verhältnisses zwischen signifikanten ( $p \leq .05$ , zweiseitig) und nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) in Abhängigkeit des Populationszusammenhangs zu veranschaulichen, sind die simulierten Effektgrößen für den jeweiligen Populationszusammenhang von  $\rho = .00$ ,  $\rho = .25$  und  $\rho = .50$  abgebildet (Abbildung 10; 11; 12).



**Abbildung 10.**  $\rho = .00$ ;  $k = 9,800$ ;  $N = 30 - 1,000$ ; Signifikanzfunktion ( $p \leq .05$ , zweiseitig).



**Abbildung 11.**  $\rho = .25$ ;  $k = 9,800$ ;  $N = 30 - 1,000$ ; Signifikanzfunktion ( $p \leq .05$ , zweiseitig).



**Abbildung 12.**  $\rho = .50$ ;  $k = 9,800$ ;  $N = 30 - 1,000$ ; Signifikanzfunktion ( $p \leq .05$ , zweiseitig).



Bei einem Populationszusammenhang von  $\rho = .50$  wären nur 1% nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) – ohne Publikationsbias – zu erwarten (Abbildung 12; Tabelle 8, Simulation 1). Selbst bei einem Populationszusammenhang von  $\rho = .25$  wären lediglich 7% nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) zu erwarten (Abbildung 11; Tabelle 8, Simulation 1). Ein Grund für die geringe Veränderung des Verhältnisses zwischen signifikanten und nichtsignifikanten Effektgrößen, scheint hierbei die Höhe der Stichprobengrößen zu sein. Ab einer Stichprobengröße von ungefähr  $n = 200$  wären bei einem Populationszusammenhang von  $\rho = .25$  ausschließlich signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) zu erwarten (Abbildung 11). Dementsprechend wäre in dem Fall zur Identifikation des Publikationsbias der Bereich der Stichproben mit  $n < 200$  von Interesse, bei denen sowohl signifikante ( $p \leq .05$ , zweiseitig) als auch nicht signifikante ( $p > .05$ , zweiseitig) Effektgrößen zu erwarten wären.

### 4.6.5 Bestimmung des Sensitivitätsbereichs zur Indexermittlung

Mit dem Software-Paket „GPower“ von Faul und Erdfelder (1992) kann unter Berücksichtigung der Teststärke, des Signifikanzniveaus und des Populationszusammenhangs, die Höhe der Stichprobengröße berechnet werden, ab wann ausschließlich signifikante Effektgrößen zu erwarten wären. Den Populationszusammenhang und die Signifikanzfunktion aus Abbildung 11 berücksichtigend ( $\rho = .25$ ;  $p \leq .05$ , zweiseitig; siehe auch Tabelle 8, Simulation 1) sowie einer Teststärke von .95, kann mit „GPower“ (Faul & Erdfelder, 1992) eine Stichprobengröße von  $n = 197$  berechnet werden, ab der ausschließlich signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) zu erwarten wären. Bei Stichprobengrößen die kleiner als  $n = 197$  wären, würden sowohl signifikante ( $p \leq .05$ , zweiseitig) als auch nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) erwartet werden können. Entsprechend wäre hier Stichproben mit Stichprobengrößen von  $n < 197$  von Interesse, um ein Publikationsbias mit Hilfe des Index zu untersuchen.

In einer entsprechend angepassten Simulation, in der die Stichproben mit einem  $n > 197$  exkludiert wurden, jedoch die restlichen Parameter aus der Studie 1 unverändert blieben, zeigte sich bei einem Populationszusammenhang von  $\rho = .25$  ein Verhältnis von 70,2% signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig;  $k = 1,263$ ) zu 39,8% nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig;  $k = 537$ ). Daraus kann nach Formel 5 ein Index von .70 berechnet werden. Bei einem unveränderten Populationszusammenhang und ohne Publikationsbias hat sich der Index um .23 bzw. 23% im Vergleich zur Simulation 1 (Index = .93), bei der Stichprobengrößen von  $n = 30$  bis  $n = 1,000$  berücksichtigt wurden, verringert. Allerdings kann es unter Umständen vorkommen, dass das oben beschriebene Vorgehen zu Fehlentscheidungen hinsichtlich der Beurteilung eines

Publikationsbias führen kann. Dazu ein Beispiel: Der Populationszusammenhang beträgt in Ackerman und Heggestad (1997) für die fluide Intelligenz und Offenheit für Erfahrungen  $\rho = .08$ . Mit dem Software-Paket von Faul und Erdfelder (1992) kann bei einem Populationszusammenhang von  $\rho = .08$  eine Stichprobengröße von  $n = 2,020$  berechnet werden, ab der ausschließlich signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) bei einer Teststärke von .95 zu erwarten wären. Dementsprechend sind bei Stichproben mit Stichprobengrößen von  $n < 2,020$  sowohl signifikante ( $p \leq .05$ , zweiseitig) als auch nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) zu erwarten. Jedoch kann angenommen werden, dass die Vergleichbarkeit des Index auf der Grundlage einer Simulation mit einer maximalen Stichprobengröße von  $n = 2,020$  und des Index aus den tatsächlichen publizierten Studien unpräzise wäre. In der Simulation würden Stichprobengrößen kontinuierlich in 10er Schritten mit jeweils 100 Stichproben bis zu einer maximalen Stichprobengröße von 2,020 erzeugt werden. Die tatsächlich publizierten Studien aus der vorliegenden Arbeit haben ebenfalls Stichproben von  $n > 1,000$ . Allerdings ist die Anzahl der Stichproben mit  $n > 1,000$  sehr viel geringer (vgl. Tabelle 7, S. 49). Dementsprechend kann davon ausgegangen werden, dass die aus der Simulation gezogenen Effektgrößen mit  $n > 1,000$  hinsichtlich der Häufigkeit nicht mit den tatsächlich publizierten Effektgrößen vergleichbar sind. Darüber hinaus erhöht sich die Wahrscheinlichkeit, dass Effektgrößen mit  $n > 1,000$  signifikant wären. Die Folge wäre ein zu strenges Kriterium, was fälschlicherweise zur Annahme eines Publikationsbias führen könnte.

Um die zuvor beschriebenen Annahmen empirisch zu überprüfen, wurden zwei weitere Simulationen mit einem Populationszusammenhang von  $\rho = .08$  durchgeführt. In der ersten Simulation wurden Effektgrößen mit Stichprobengrößen von  $n = 30$  bis  $n = 2,020$  erzeugt. Dabei waren 63% der gezogenen Effektgrößen signifikant ( $p \leq .05$ , zweiseitig). In einer zweiten Simulation wurden Effektgrößen mit Stichprobengrößen von  $n = 30$  bis  $n = 900$  erzeugt. In dem Fall waren lediglich 33.37% der gezogenen Effektgrößen signifikant ( $p \leq .05$ , zweiseitig). Der prozentuale Anteil signifikanter Effektgrößen ( $p \leq .05$ , zweiseitig) hat sich – bei unverändertem Populationszusammenhang und ohne Publikationsbias – in der zweiten Simulation um 29.63% (63% minus 33.37%) verringert. Dementsprechend sollte der Erzeugung von Effektgrößen mit Stichprobengrößen, die in den real publizierten Studien in der Häufigkeit und Höhe nicht vorhanden sind, vorgebeugt werden, um eine reelle Vergleichbarkeit der Indizes hinsichtlich eines Publikationsbias zu gewährleisten.

#### 4.6.7 Adaptation der Simulation hinsichtlich der Stichprobengrößen und –anzahl real publizierter Studien

Die simulierten Effektgrößen an der Höhe und Häufigkeit der real publizierten Stichprobengrößen auszurichten, scheint angemessen zu sein, um eine präzise Vergleichbarkeit der Indizes bezüglich eines Publikationsbias zu ermöglichen. Dies macht eine weitere Adaptation der Simulation notwendig, die im Folgenden beschrieben werden soll.

Bisher wurden in den Simulationen die Stichprobengrößen in 10er Schritten erhöht (bspw.  $n = 30$ , 40, 50, usw.) und 100 Effektgrößen je Stichprobengröße gezogen (siehe oben). Um die Höhe und Häufigkeit der Stichprobengrößen real publizierter Studien in einer Simulation zu berücksichtigen, wurden Intervalle von Stichprobengrößen mit einer Spannweite von  $n = 40$  gebildet. Die Intervalle wurden bis zu einer maximalen Stichprobengröße, ab der nur noch signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) zu erwarten sind („GPower“; Faul & Erdfelder, 1992), berücksichtigt. Zur Veranschaulichung:

- 1. Intervall  $n = 10$  bis  $n = 50$ ,
- 2. Intervall  $n = 60$  bis  $n = 100$ ,
- 3. Intervall  $n = 110$  bis  $n = 150$ ,

bis zu dem letzten Intervall  $n_i$ , dass die Stichprobengrößen einschließt, für die nicht ausschließlich signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) zu erwarten sind.

In einem nächsten Schritt wurden die Stichprobengrößen der real publizierten Studien den entsprechenden Intervallen zugeordnet. Anschließend wurde die Anzahl der Studien in dem jeweiligen Intervall summiert. Aus den summierten Werten wurde für das jeweilige Intervall der prozentuale Anteil aus den gesamten publizierten Studien berechnet. Anhand der prozentualen Verteilungen der Stichprobengrößen real publizierter Studien je Intervall, wurde die Simulation adaptiert. Ein fiktives Beispiel soll dies veranschaulichen: Der Populationszusammenhang beträgt  $\rho = .14$  und die Stichprobengrößen sollen eine Spannweite von  $n = 15$  bis  $n = 473$  haben. Mit dem Software-Paket von Faul und Erdfelder (1992) wurde bei einem Populationszusammenhang von  $\rho = .14$  und einer Teststärke von .95 eine Stichprobengröße von  $n = 652$  berechnet, ab der ausschließlich signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) zu erwarten wären. Dementsprechend wurden hier alle publizierten Studien inkludiert, da die maximale Stichprobengröße von  $n = 473$  kleiner als der berechnete Grenzwert von  $n = 652$  ist. Die Anzahl der (fiktiven) publizierten Studien wurde unter Berücksichtigung der Stichprobengröße den Intervallen zugeordnet und entsprechend die prozentuale Verteilung gebildet (Tabelle 9).

Anschließend wurde eine Simulation unter Berücksichtigung der prozentualen Verteilung der Stichprobengrößen je Intervall aus den (fiktiv) publizierten Studien durchgeführt.

**Tabelle 9**

Fiktives Beispiel publizierter Stichproben, die den Intervallen nach Stichprobengrößen zugeordnet wurden und anschließend adaptierter Simulation

| Intervall | $N$       | $k$ (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | $k$ (Simulation) <sup>3</sup> |
|-----------|-----------|------------------------------|-----------------------|-------------------------------|
| 1.        | 0 – 50    | 10                           | 8.06                  | 403.0                         |
| 2.        | 60 – 100  | 24                           | 19.35                 | 967.5                         |
| 3.        | 110 – 150 | 28                           | 22.58                 | 1,129.0                       |
| 4.        | 160 – 200 | 20                           | 16.13                 | 806.0                         |
| 5.        | 210 – 250 | 16                           | 12.90                 | 645.0                         |
| 6.        | 260 – 300 | 13                           | 10.48                 | 524.0                         |
| 7.        | 310 – 350 | 6                            | 4.84                  | 242.0                         |
| 8.        | 360 – 400 | 4                            | 3.23                  | 161.5                         |
| 9.        | 410 – 450 | 1                            | 0.81                  | 40.5                          |
| 10.       | 460 – 500 | 2                            | 1.63                  | 81.5                          |
| $\Sigma$  |           | <b>124</b>                   | <b>100.00</b>         | <b>5,000.0</b>                |

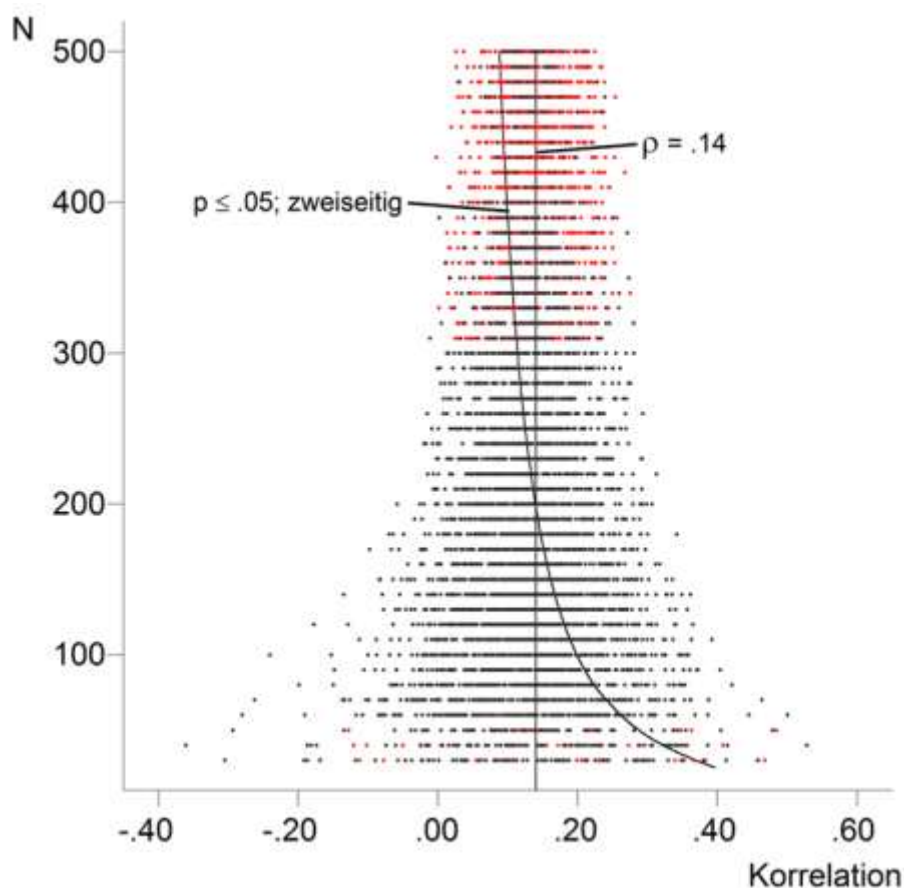
Anmerkungen.  $\rho = .14$ ; <sup>1</sup> = Anzahl (fiktiv) publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten  $k$  für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt  $k$  von 5,000 gezogenen Effektgrößen.

Um in der Simulation eine ausreichend große Anzahl von Effektgrößen zu ziehen, wurde ein  $k$  von 5,000 (= 100%) festgelegt. Somit werden bereits bei einer 1%igen Gewichtung 50 Effektgrößen aus den generierten Zufallsstichproben gezogen. Um die Zielgröße von  $k = 5,000$  zu realisieren, wurde eine zusätzliche Multiplikation der prozentualen Gewichtung mit 10 durchgeführt. Dazu folgendes Beispiel aus der Tabelle 9: Im dritten Intervall ( $n = 110$  bis  $n = 150$ ) beträgt die Zahl der (fiktiven) publizierten Studien  $k = 28$  bei einer Gesamtanzahl von  $k = 124$ . Der prozentuale Anteil im dritten Intervall beträgt damit 22.58%. Der Wert 22.58 wurde mit 10 multipliziert, woraus sich 225.8 berechnet. Nachfolgend wurden jeweils 225.8 (gerundet 226) Effektgrößen für  $n = 110$ ,  $n = 120$ ,  $n = 130$ ,  $n = 140$  und  $n = 150$  gezogen. Insgesamt wurden somit 1,129 Effektgrößen im dritten Intervall gezogen (Tabelle 9). Von den 1,129 Effektgrößen waren bei einem Populationszusammenhang von  $\rho = .14$  – ohne Publikationsbias – 29.6% Effektgrößen signifikant ( $p \leq .05$ , zweiseitig) und 70.4% Effektgrößen nichtsignifikant ( $p > .05$ , zweiseitig; Index .30).

Insgesamt kann unter den gegebenen Parametern und mit Hilfe der adaptierten Simulation (Tabelle 9) ein prozentuales Verhältnis von 38% signifikanten ( $p \leq .05$ , zweiseitig) zu 62% nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) – ohne Publikationsbias – erwartet werden. Zum Vergleich: In einer zusätzlichen durchgeführten Simulation, in der keine prozentuale

Gewichtung der Stichprobengrößen anhand der (fiktiven) publizierten Studien aus Tabelle 9 erfolgte, die restlichen Parameter jedoch unverändert blieben ( $\rho = .14$ ;  $k = 5,000$ ; Stufung der Stichprobengrößen in 10er Schritten von  $n = 10$  bis  $n = 500$ ), waren bereits 52% der Effektgrößen – ohne Publikationsbias – signifikant ( $p \leq .05$ , zweiseitig). Ohne die Berücksichtigung der Gewichtung von Stichprobengrößen publizierter Studien, hat sich in der Simulation der prozentuale Anteil signifikanter Effektgrößen um 14%, von 38% auf 52%, erhöht.

In Abbildung 13 sind die beiden zuvor beschriebenen Simulationen abgebildet. Die Verteilung der Effektgrößen aus der Simulation, bei der eine prozentuale Gewichtung der Stichprobengrößen anhand der (fiktiven) publizierten Studien erfolgte (Tabelle 9) ist in Schwarz abgebildet. Die Verteilung der Effektgrößen aus der zweiten Simulation, in der keine prozentuale Gewichtung der Stichprobengrößen anhand der (fiktiven) publizierten Studien aus Tabelle 9 erfolgte, ist eine Ebene dahinter, in Rot abgebildet.



**Abbildung 13**  $\rho = .14$ ; eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig); Simulation 1: Effektgrößen unter Berücksichtigung der prozentual gewichteten Stichprobengrößen (schwarze Punkte); Simulation 2: Effektgrößen ohne prozentuale Gewichtung publizierter Stichprobengrößen (rote Punkte); beide Simulationen  $k = 5,000$ ; Simulation 2 ist eine Ebene unter Simulation 1 abgebildet.

## Kapitel 4 – Methode

In dem Bereich der Stichprobengrößen von  $n = 10$  bis  $n = 300$  befinden sich 89.5% der gezogenen Effektgrößen ( $k = 4,475$ ) aus der adaptierten Simulation (in Schwarz, Abbildung 13). Hingegen befinden sich lediglich 60% der gezogenen Effektgrößen ( $n = 3,000$ ) aus der Simulation, in der keine Adaptation der prozentualen Gewichtung der Stichprobengrößen aus den (fiktiv) publizierten Studien erfolgte, in dem Bereich der Stichprobengrößen von  $n = 10$  bis  $n = 300$ . Ab einer Stichprobengröße von  $n = 300$  werden mehr Effektgrößen aus der Simulation gezogen, bei der keine prozentuale Gewichtung der Stichprobengrößen aus den (fiktiv) publizierten Studien erfolgte als bei der adaptierten Simulation (Abbildung 13).

Der Schnittpunkt der eingezeichneten Signifikanzfunktion ( $p \leq .05$ , zweiseitig) und des eingezeichneten Populationszusammenhangs von  $\rho = .14$  liegt bei einer Stichprobengröße von  $n = 200$  (Abbildung 13). Unterhalb dieser Stichprobengröße von  $n = 200$  würden in dem Fall vermehrt nichtsignifikante Effektgrößen ( $p \leq .05$ , zweiseitig) in einer Simulation gezogen werden und oberhalb dieser Stichprobengröße würden häufiger signifikante ( $p \leq .05$ , zweiseitig) als nichtsignifikante Effektgrößen ( $p < .05$ , zweiseitig) gezogen werden (Abbildung 13). In der Simulation, in der keine Adaptation der prozentualen Gewichtung der Stichprobengrößen aus den (fiktiv) publizierten Studien vorgenommen wurde, wurden die Effektgrößen konstant 100-mal je Stichprobengröße von  $n = 10$  bis  $n = 500$  in 10er Schritten gezogen. Das bedeutet, dass bis zu einer einschließlichen Stichprobengröße von  $n = 200$  erst 40% der gesamten Effektgrößen gezogen wurden. 60% der gezogenen Effektgrößen befinden sich in dem Bereich der Stichprobengrößen von  $n = 210$  bis  $n = 500$ , in dem häufiger signifikante ( $p \leq .05$ , zweiseitig) als nichtsignifikante Effektgrößen ( $p < .05$ , zweiseitig) aufgrund der Stichprobengröße gezogen werden. Dementsprechend konnten in der nichtadaptierten Simulation insgesamt mehr signifikante ( $p \leq .05$ , zweiseitig) als nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) – ohne Publikationsbias – erwartet werden. Anders bei der adaptierten Simulation: Hierbei liegen 66.12% der Effektgrößen in dem Bereich von  $n \leq 200$ . Entsprechend konnten in der adaptierten Simulation insgesamt mehr nichtsignifikante ( $p > .05$ , zweiseitig) als signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) – ohne Publikationsbias – erwartet werden.

Der Nachteil bezüglich eines Vergleichs der Verteilung mit real publizierten Effektgrößen zeigt sich in der Simulation, in der keine prozentuale Gewichtung der Stichprobengrößen aus den (fiktiv) publizierten Studien berücksichtigt wurde: Es werden Effektgrößen gezogen, denen Stichprobengrößen zugrundeliegen, die in der Höhe und Häufigkeit real publizierter Studien nicht vorhanden sind (Abbildung 13, Bereich der Stichprobengrößen:  $n = 300$  bis  $n = 500$ , rot

dargestellte Effektgrößen). Die Folge ist ein strengeres Kriterium in Form eines höheren Index. Die Zunahme des Index um .14 (= 14%) steht nicht im Verhältnis zu einem möglichen Publikationsbias, sondern ist die Folge einer vermehrten Ziehung von signifikanten Effektgrößen mit großem  $n$ .

Um den Einfluss der Stichprobengrößen auf die simulierte Verteilung der signifikanten ( $p \leq .05$ , zweiseitig) und nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zu kontrollieren, wurden die Effektgrößen aus der Simulation entsprechend der prozentualen Gewichtung der tatsächlich publizierten Stichprobengrößen gezogen. Dadurch wurden die publizierten Verteilungen von Effektgrößen, bei denen die Möglichkeit eines Publikationsbias besteht, gezielt mit den simulierten Verteilungen – ohne Publikationsbias – verglichen.

Darüber hinaus gilt es eine Besonderheit hinsichtlich der Mindestanzahl tatsächlich publizierter Studien zu berücksichtigen: Die Anzahl der empirischen Effektgrößen begrenzt die Genauigkeit der Simulation. Dabei ist der extremste mögliche Wert des empirischen Indexes durch  $1/k$  begrenzt. Das bedeutet, dass in den Fällen, in denen  $k$  sehr klein ist, die Feinheit der Simulation möglicherweise keinen Sinn macht bzw. Bereiche durch die Simulation exploriert werden, die kleiner als  $1/k$  der empirischen Studien sind und dementsprechend fälschlicherweise einen Publikationsbias anzeigen. Um eine adäquate Vergleichbarkeit des empirischen Index mit dem aus der Simulation ermittelten Index zu gewährleisten, wurde in den Fällen, in denen der Index aus der Simulation kleiner war als der extremste mögliche Wert des empirischen Index ( $1/k$ ), der Index aus der Simulation auf  $1/k$  der empirischen Studien aufgerundet.

#### **4.6.8 Verwendete Methoden zur Untersuchung des Publikationsbias in der vorliegenden Arbeit**

Untersuchungen zur Identifikation des Publikationsbias berücksichtigen eine asymmetrische Verteilung publizierter Effektgrößen (bspw. Duval & Tweedie, 2000a, 2000b; Light & Pillemer, 1984) und/oder die vorzugsweise Publikation signifikanter Effektgrößen (bspw. Hedges & Olkin, 1985; Rosenthal, 1979). In der vorliegenden Arbeit wurde die bereits etablierte „trim and fill“-Methode (Duval & Tweedie, 2000a, 2000b) angewendet, nach der von einem Publikationsbias auszugehen ist, wenn sich die publizierten Effektgrößen asymmetrisch um die mittlere Effektgröße verteilen (siehe Kapitel 4.6.1). Darüber hinaus wurde eine im Rahmen der vorliegenden Arbeit entwickelte Methode angewendet, die die Tendenz berücksichtigt, dass eher signifikante als nichtsignifikante Effektgrößen zur Publikation gelangen (vgl. Ferguson & Heene, 2012; Rosenthal,

1979). Dazu wurde das Verhältnis zwischen publizierten signifikanten ( $p \leq .05$ , zweiseitig) und nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) untersucht, in dem die Differenz der Indizes (Formel 10) zwischen den tatsächlich publizierten Effektgrößen – mit einem möglichen Publikationsbias – und den simulierten Effektgrößen – ohne Publikationsbias – gebildet wurde (vgl. Kapitel 4.6.4 – 4.6.7). Wichen die Indizes voneinander ab, wurde anschließend berichtet, wieviel nichtsignifikante Effektgrößen ( $p \leq .05$ , zweiseitig) zusätzlich publiziert sein müssten, um ein Publikationsbias ausschließen zu können.

Die Simulationsmethode wurde im Rahmen der vorliegenden Arbeit entwickelt. Um die Komplexität nicht noch weiter zu erhöhen, wurde bspw. bei der Indexberechnung der tatsächlich publizierten Studien keine Gewichtung der Stichprobengrößen berücksichtigt. Traten sehr große bzw. heterogene Stichproben auf, führte dies dazu, dass in einigen Fällen keine schlüssige Interpretation bezüglich eines Publikationsbias möglich war. In diesen Fällen wurde auf eine Ergebnisdarstellung zum Publikationsbias mittels Simulationsmethode verzichtet (siehe auch Kapitel 6.3 Grenzen der vorliegenden Arbeit und verbleibendes Arbeitsprogramm).



## 5 Ergebnisse

Die Ergebnisse zum Zusammenhang mit der Allgemeinen Intelligenz (g), Fluiden Intelligenz (Gf) und kristallisierten Intelligenz (Gc) werden nacheinander für die Persönlichkeitsdimensionen Neurotizismus (Ne), Extraversion (E), Offenheit für Erfahrung (O), Verträglichkeit (V) und Gewissenhaftigkeit (G) berichtet. In der jeweils ersten metaanalytischen Untersuchung werden die Befunde, bei denen eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien erfolgte, berichtet. Danach werden jeweils die Ergebnisse berichtet, denen eine Zuordnung der Fähigkeitskonstrukte durch den Autor der vorliegenden Arbeit nach der CHC-Theorie (McGrew, 2009) zugrunde liegt. Anschließend werden jeweils die metaanalytisch gerechneten Varianten (Zuordnung der Fähigkeitskonstrukte gemäß Autoren vs. CHC) hinsichtlich eines Publikationsbias untersucht. Bei einer erklärten Varianz von weniger als 75% werden, der Empfehlung von Hunter und Schmidt (2004) folgend, jeweils Moderatoranalysen durchgeführt.

### 5.1 Ergebnisse zur Allgemeinen Intelligenz und Neurotizismus

*Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien:*

In der Variante beträgt die inkludierte Studienanzahl  $k = 69$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{\text{bare-bones}} = -.11$  ( $SD = .04$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 1,332$  (Range:  $n = 59$  bis  $71,887$ ;  $SD = 8,625.76$ ;  $N = 91,933$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. –überschätzung korrigierte Effektgröße beträgt  $\rho = -.13$  ( $SD_{\rho} = .06$ ). Die Null liegt hierbei nicht im Glaubwürdigkeitsintervall („credibility interval“). Die erklärte Varianz beträgt 26.90%. (Tabelle 10).

Dem Vorschlag von Hunter und Schmidt (2004) folgend, wurde anschließend die Studie von Reeve et al. (2006) mit einer Stichprobengröße von  $n = 71,887$  exkludiert, um den Einfluss einer einzigen Studie mit einer sehr großen Stichprobengröße zu kontrollieren. Der Betrag der unkorrigierten Effektgröße verringerte sich von  $\bar{r}_{\text{bare-bones}} = -.11$  ( $SD = .04$ ) auf  $\bar{r}_{\text{bare-bones}} = -.09$  ( $SD = .07$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 295$  (Range:  $n = 59$  bis  $1,912$ ;  $SD = 351.73$ ,  $N = 20,046$ ). Der Betrag der artefaktkorrigierten Effektgröße verringert sich hierbei von  $\rho = -.13$  ( $SD_{\rho} = .06$ ) auf  $\rho = -.09$  ( $SD_{\rho} = .10$ ). Die Null liegt im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 32.19% (Tabelle 10).

Ohne die Exklusion der Studie von Reeve et al. (2006) kann ein negativer Zusammenhang zwischen g und Ne angenommen werden, da die Null nicht im Glaubwürdigkeitsintervall liegt. Jedoch sollte die Höhe der artefaktkorrigierten Effektgröße nicht verallgemeinert werden, da die

erklärte Varianz deutlich kleiner als 75% ist. Es ist davon auszugehen, dass die Höhe der artefaktkorrigierten Effektgröße von Moderatoren abhängt. Dementsprechend wurden Moderatoranalysen durchgeführt (siehe unten). Durch die Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) verändert sich der metaanalytische Befund bezüglich der Generalisierbarkeit substantiell. Das Glaubwürdigkeitsintervall schließt die Null mit ein und die erklärte Varianz ist auch hierbei deutlich kleiner als 75%. Nach Kemery et al. (1987) kann bei einem derartigen Befund angenommen werden, dass ein mittlerer Effekt heterogener Subpopulationen vorliegt, bei der die Höhe sowie die Richtung der Effektgröße nicht verallgemeinert werden sollte. Auch in diesen Fall wurden Moderatoranalysen durchgeführt.

**Tabelle 10**

Metaanalytische Ergebnisse zum Zusammenhang zwischen g und Ne

| g/Ne                                 | $N$    | $k$ | $\rho$ | $SD_{\rho}$ | %Var. | 90% CV | $\bar{r}$ | $SD_r$ | %Var  |
|--------------------------------------|--------|-----|--------|-------------|-------|--------|-----------|--------|-------|
| kodiert nach Autor(en) <sup>1</sup>  | 91,933 | 69  | -.13   | .059        | 26.90 | -.06   | -.11      | .036   | 36.43 |
| kodiert nach Autor(en) <sup>1a</sup> | 20,046 | 68  | -.09   | .100        | 32.19 | .03    | -.09      | .074   | 37.80 |
| kodiert nach CHC <sup>2</sup>        | 92,425 | 76  | -.14   | .045        | 42.28 | -.08   | -.11      | .031   | 45.01 |
| kodiert nach CHC <sup>2a</sup>       | 20,538 | 75  | -.11   | .082        | 46.17 | -.01   | -.09      | .064   | 46.77 |

*Anmerkungen.*  $N$  = Stichprobengröße;  $k$  = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und –überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; %Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall („credibility interval“), aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analyse); <sup>1</sup> = Datensätze, in denen eine Zuordnung von g gemäß der Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von g gemäß der Autoren der publizierten Studien vorgenommen wurde und die Studie von Reeve et al. (2006,  $n = 71,887$ ) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein nach der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein nach der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde und die Studie von Reeve et al. (2006,  $n = 71,887$ ) exkludiert ist.

*Zuordnung der Fähigkeitskonstrukte gemäß CHC-Theorie:*

In dieser Variante beträgt die inkludierte Studienzahl  $k = 76^6$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = -.11$  ( $SD = .03$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 1,216$  (Range:  $n = 57$  bis  $71,887$ ;  $SD = 8,219.60$ ;  $N = 92,425$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. –überschätzung korrigierte Effektgröße beträgt  $\rho = -.14$  ( $SD_{\rho} = .05$ ). Die erklärte Varianz beträgt 42.28% und die Null liegt

<sup>6</sup> Wenn das Fähigkeitskonstrukt g explizit in den publizierten Studien berichtet ist, wurde geprüft, ob die Zuordnung des Tests zum Fähigkeitskonstrukt gemäß der CHC-Theorie (McGrew, 2009) als g beibehalten werden konnte. Wenn g nicht explizit in den publizierten Studien berichtet ist, wurde geprüft, ob eine Zuordnung des jeweiligen Tests zu g nach der CHC-Theorie (McGrew, 2009) möglich war. Dadurch ist die inkludierte Studienzahl in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie erfolgte, größer ist als in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde. Siehe dazu auch Studieninklusion und Zuordnung in Kapitel 4.2 und Tabelle 6.

## Kapitel 5 – Ergebnisse

nicht im Glaubwürdigkeitsintervall. Auch in dieser metaanalytischen Untersuchung wurde eine Variante gerechnet, in der die Studie von Reeve et al. (2006,  $n = 71,887$ ) exkludiert wurde. Nach der anschließenden Exklusion der Studie von Reeve et al. (2006) verringerte sich der Betrag der unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = -.11$  ( $SD = .03$ ) auf  $\bar{r}_{bare-bones} = -.09$  ( $SD = .06$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 273$  (Range:  $n = 57$  bis  $1,912$ ;  $SD = 288.99$ ;  $N = 20,538$ ). Der Betrag der artefaktkorrigierten Effektgröße verringert sich hierbei von  $\rho = -.14$  ( $SD_{\rho} = .05$ ) auf  $\rho = -.11$  ( $SD_{\rho} = .08$ ). Das Glaubwürdigkeitsintervall schließt die Null nicht ein und die erklärte Varianz beträgt 46.17% (Tabelle 10).

Unabhängig von der Exklusion der Studie von Reeve et al. (2006,  $n = 71, 887$ ) kann von einem negativen Zusammenhang zwischen  $g$  und  $N_e$  ausgegangen werden, da das Glaubwürdigkeitsintervall die Null nicht einschließt. Da allerdings die Varianzerklärung deutlich kleiner als 75% ist, kann angenommen werden, dass die Höhe der artefaktkorrigierten Effektgröße von Moderatoren abhängt. Dementsprechend wurden hier ebenfalls Moderatoranalysen durchgeführt (siehe unten).

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde:*

Bei den im Ergebnisteil durchgeführten Untersuchungen zum Publikationsbias wurden die Studien von Reeve et al. (2006) und De Fruyt et al. (2006) mit Stichprobengrößen von  $n = 71,887$  bzw.  $n = 20,175$  nicht berücksichtigt (siehe Kapitel 4.4; 4.6.4). Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe der Schätzers mit  $R_0 < 1$  eine weitestgehend symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.07$  angenommen werden. Der zweite Schätzer zeigt mit  $L_0 = 1.46$  eine asymmetrische Verteilung an. Nach der Exklusion der zwei Effektgrößen ( $L_0^+ = 2$ ) mit negativen Vorzeichen und den höchsten Beträgen kann eine symmetrische Verteilung der publizierten Effektgrößen ( $k = 66$ ) um die unveränderte mittlere Effektgröße von  $\bar{r} = -.07$  mit  $L_0 < 1$  angenommen werden.

Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 65 relevanten Stichprobengrößen real publizierter Studien mit  $n < 1,594$  und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 11 aufgeführt.

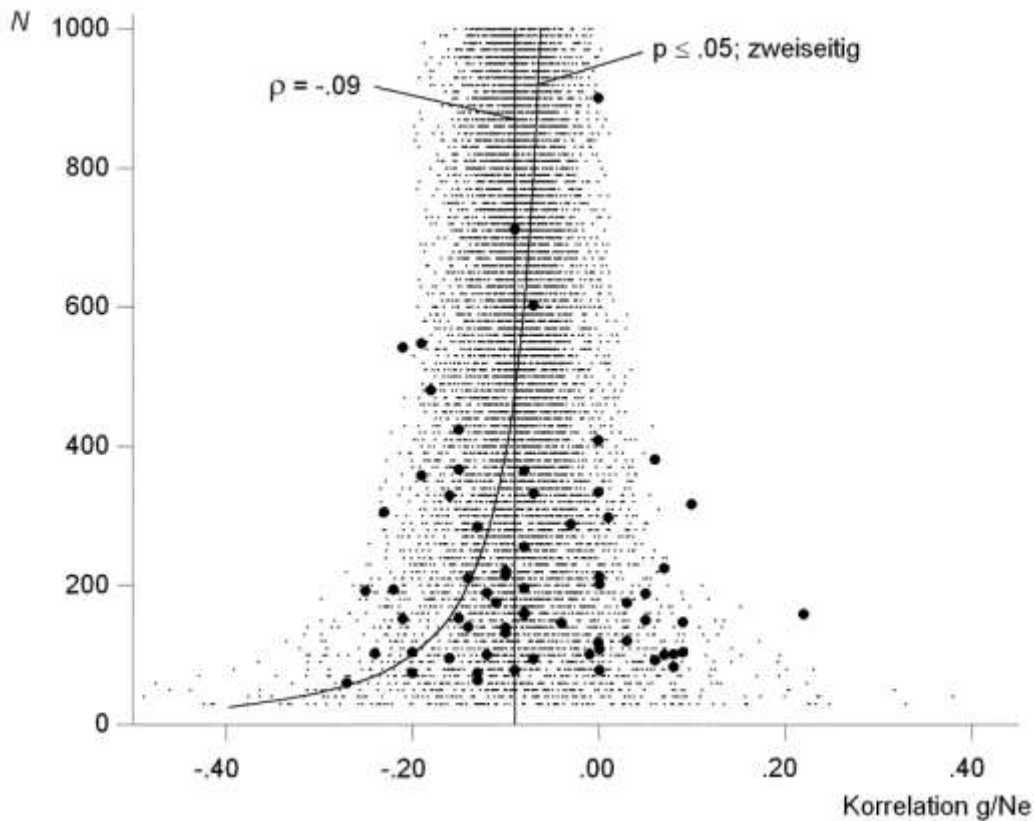
**Tabelle 11**

Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | <i>N</i>      | <i>k</i> (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | <i>k</i> (Simulation) <sup>3</sup> |
|-----------|---------------|-----------------------------------|-----------------------|------------------------------------|
| 1.        | 60 – 100      | 17                                | 25.76                 | 1,288                              |
| 2.        | 110 – 150     | 11                                | 16.67                 | 834                                |
| 3.        | 160 – 200     | 11                                | 16.67                 | 834                                |
| 4.        | 210 – 250     | 5                                 | 7.58                  | 379                                |
| 5.        | 260 – 300     | 4                                 | 6.06                  | 303                                |
| 6.        | 310 – 350     | 5                                 | 7.58                  | 379                                |
| 7.        | 360 – 400     | 4                                 | 6.06                  | 303                                |
| 8.        | 410 – 450     | 2                                 | 3.03                  | 152                                |
| 9.        | 460 – 500     | 1                                 | 1.52                  | 76                                 |
| 10.       | 510 – 550     | 2                                 | 3.03                  | 152                                |
| 11.       | 560 – 600     | 1                                 | 1.52                  | 76                                 |
| 12.       | 710 – 750     | 1                                 | 1.52                  | 76                                 |
| 13.       | 860 – 900     | 1                                 | 1.52                  | 76                                 |
| 14.       | 1,310 – 1,350 | 1                                 | 1.52                  | 76                                 |
| <b>Σ</b>  |               | <b>66</b>                         | <b>100.00</b>         | <b>5,004<sup>4</sup></b>           |

Anmerkungen.  $\rho = -.11$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten *k* für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt *k* von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt *k* = 5,004.

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .23 hat sich aus einem Verhältnis von 1,142 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 3,862 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Da die simulierten Effektgrößen keinerlei Publikationsbias aufweisen, kann hierbei ein Verhältnis von 23% signifikanten ( $p \leq .05$ , zweiseitig) zu 77% nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) als Referenz zur Beurteilung eines Publikationsbias herangezogen werden. Der aus den tatsächlich publizierten Effektgrößen berechnete Index beträgt .27 (18 signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 48 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 14).



**Abbildung 14.**  $\rho = -.09$ ; tatsächlich publizierte Effektgrößen als schwarz ausgefüllte Kreise abgebildet; gezogene Effektgrößen aus der Simulation als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 11) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien.

Um bei den tatsächlich publizierten Effektgrößen einen Index von .23 – ohne Publikationsbias – zu erreichen, müssten 11 weitere nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von 18 signifikanten ( $p \leq .05$ , zweiseitig) zu 59 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) unter den gegebenen Parametern publiziert sein. Dementsprechend kann ein Publikationsbias nicht ausgeschlossen werden. Offenbar sind mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde.

Anhand der Abbildung 14 ist auffällig, dass im rechten Bereich zwei Effektgrößen, eindeutig außerhalb der simulierten Effektgrößen, publiziert sind. In Tabelle 12 sind zu den zwei Effektgrößen studienspezifische Merkmale aufgeführt. Allerdings ließen sich keine eindeutigen Auffälligkeiten bzw. Gemeinsamkeiten feststellen, weshalb die publizierten Effektgrößen von den simulierten Effektgrößen im rechten äußeren Bereich abweichen.

**Tabelle 12**

Deskriptive Angaben publizierter Studien, in denen die Effektgröße des Zusammenhangs zwischen  $g$  und  $Ne$  von den gezogenen Effektgrößen aus der adaptierten Simulation abweicht

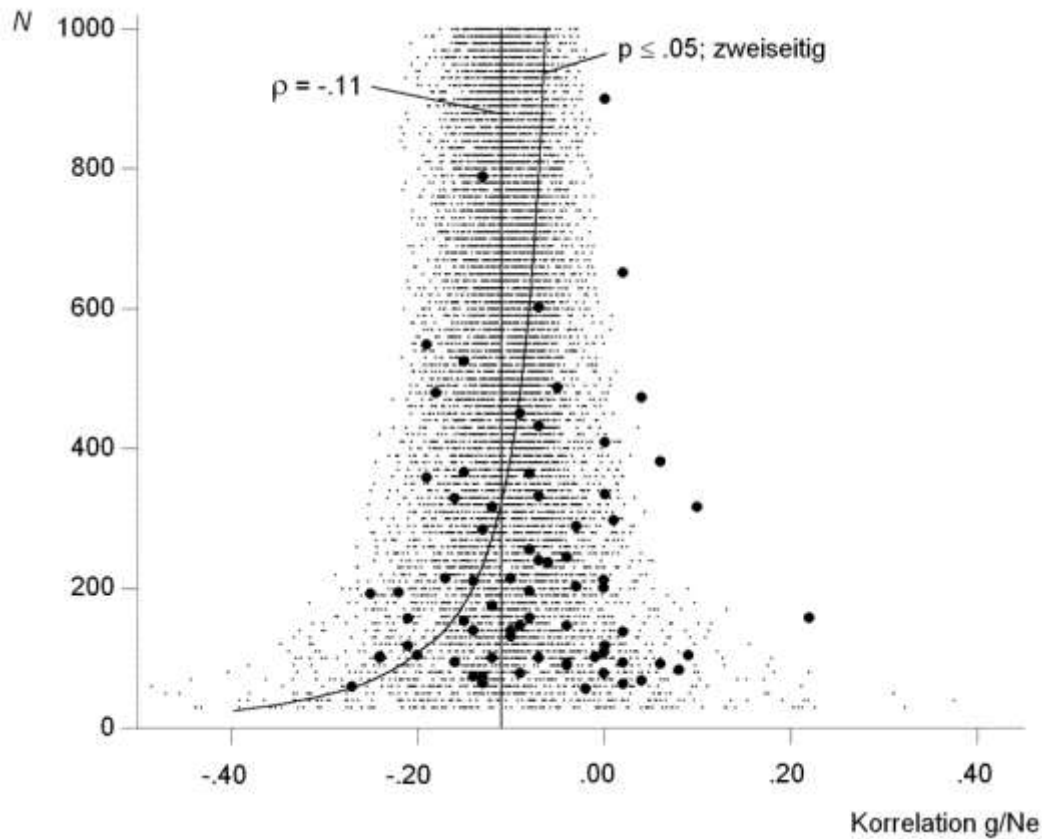
| Studie                                      | kogLT | PSK      | Land | Alter                          | Stichprobe                               | Thema             |
|---|-------|----------|------|--------------------------------|--|-------------------|
| Batey, Chamorro-Premuzic und Furnham (2010) | WPT   | NEO-PI-R | GB   | $M = 20.43$<br>( $SD = 2.45$ ) | Studenten<br>( $n = 158$ )               | Kreativität       |
| Neuman und Wright (1999)                    | TTA   | NEO-PI-R | USA  | nicht<br>spezifiziert          | Personal-<br>management<br>( $n = 316$ ) | Team Effektivität |

*Anmerkungen.* kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; WPT = Wonderlic Personal Test; TTA = Thurstone Test of Alertness.

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durchgeführt wurde:*

Nach der „trim and fill“ – Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann aufgrund der Höhe der Schätzer mit  $R_0 < 1$  und  $L_0 < 1$  eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.07$  angenommen werden.

Die Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens erlaubt keine schlüssige Interpretation der Ergebnisse hinsichtlich eines Publikationsbias. Dementsprechend wurde in dem Fall auf die Simulationsmethode zur Identifikation eines Publikationsbias verzichtet.



**Abbildung 15.**  $\rho = -.11$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet,  $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4; eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie (McGrew, 2009).

Sechs der tatsächlich publizierten Effektgrößen liegen im rechten Bereich, außerhalb der simulierten Effektgrößen (Abbildung 15). Hierbei kann der Eindruck entstehen, dass diese sechs Effektgrößen aus einer anderen Grundgesamtheit stammen als die ermittelten Effektstärken der Simulation. Allerdings lassen sich keine eindeutigen Auffälligkeiten bzw. Gemeinsamkeiten feststellen, weshalb diese sechs Studien Effektgrößen berichten, die außerhalb der simulierten Effektgrößen liegen (Tabelle 13).

**Tabelle 13**

Deskriptive Angaben publizierter Studien, in denen die Effektgröße des Zusammenhangs zwischen  $g$  und  $Ne$  von den gezogenen Effektgrößen aus der adaptierten Simulation abweicht

| Studie                                      | kogLT                | PSK      | Land    | Alter                          | Stichprobe                               | Thema   |
|---|----------------------|----------|---------|--------------------------------|--|---|
| Batey, Chamorro-Premuzic und Furnham (2010) | WPT                  | NEO-PI-R | GB      | $M = 20.43$<br>( $SD = 2.45$ ) | Studenten<br>( $n = 158$ )               | Kreativität   |
| Neuman und Wright (1999)                    | TTA                  | NEO-PI-R | USA     | nicht<br>spezifiziert          | Personal-<br>Management<br>( $n = 316$ ) | Team Effektivität   |
| Allik und Realo (1997)                      | unbekannter<br>kogLT | NEO-PI   | Estland | $M = 19.10$<br>( $SD = 2.30$ ) | angehende<br>Studenten<br>( $n = 381$ )  | akademische<br>Fähigkeiten                                  |
| Chamorro-Premuzic und Arteche (2008)        | Raven und<br>MHV     | NEO-PI-R | GB      | $M = 20.31$<br>( $SD = 3.43$ ) | Studenten<br>( $n = 473$ )               | intellektuelle<br>Kompetenzen und<br>akademischer<br>Erfolg |
| Barrick, Stewart, Neubert und Mount (1998)  | WPT                  | PCI      | USA     | $M = 41.66$<br>( $SD = 9.59$ ) | Angestellte<br>( $n = 652$ )             | Team Prozesse<br>und Effektivität                           |
| Moutafi, Furnham und Crump (2003)           | WGCTA<br>und GMA     | NEO-PI-R | GB      | $M = 44.00$<br>( $SD = n.b.$ ) | Angestellte<br>( $n = 900$ )             | Vorhersage von<br>Intelligenz durch<br>Persönlichkeit       |

*Anmerkungen.* kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; TTA = Thurstone Test of Alertness, MHV = Mill Hill Vocubular; WGCTA = Watson Glaser Critical Thinking Appraisal; GMA = Graduate and Managerial Assessment; PCI = Personel Characteristics Inventory.

*Moderatoranalysen:*

Zum Zusammenhang zwischen  $g$  und  $Ne$  beträgt die Varianz des artefaktkorrigierten Effektes ( $\rho$ ) ausnahmslos weniger als 75% (Tabelle 10). Dementsprechend wurden Moderatoranalysen durchgeführt. Die kognitiven Leistungstests der publizierten Studien berücksichtigend, wurden die Studien, in denen der Wonderlic Personnel Test (WPT; Wonderlic, 1992) zur Erfassung der allgemeinen Intelligenzausprägung eingesetzt wurde mit den Studien verglichen, in denen andere kogLT als der WPT eingesetzt wurde (bspw. Wechsler Adult Intelligence Scale, Wechsler, 1981). Des Weiteren wurden Moderatoranalysen hinsichtlich der eingesetzten Persönlichkeitsinventare durchgeführt (Tabelle 14). Dabei wurden folgende Persönlichkeitsinventare verglichen:

- NEO-Inventare (NEO-PI-(R) oder NEO-FFI) vs.
- NEO-PI-(R) vs.
- NEO-FFI vs.
- FFM-Inventare, allerdings keine NEO-Inventare (bspw. International Personality Item Pool, Personel Characteristics Inventory, Vijf Persoonlijkheids-Factoren Test, Perspectives Inventory) vs.
- Eysenck-Inventare (bspw. Eysenck Personality Questionnaire, Eysenck Personality Profiler).



In Tabelle 14 sind die metaanalytischen Kennwerte der entsprechenden Moderatoranalysen zum Zusammenhang zwischen g und Ne aufgeführt. Dabei wurden die Moderatoranalysen separat nach der Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien und der CHC-Theorie (McGrew, 2009) aufgeführt.

**Tabelle 14**  
Moderatoranalysen zum Zusammenhang zwischen g und Ne

| g/Ne                    | Moderator   | N      | k  | $\rho$ | $SD_{\rho}$ | % Var. | 90% CV | $\bar{r}$ | $SD_r$ | % Var  |
|-------------------------|-------------|--------|----|--------|-------------|--------|--------|-----------|--------|--------|
| <i>kogLT</i>            |             |        |    |        |             |        |        |           |        |        |
| Autor(en) <sup>1</sup>  | WPT         | 4,601  | 26 | -.09   | .099        | 44.66  | .04    | -.07      | .082   | 45.80  |
| Autor(en) <sup>1</sup>  | ohne WPT    | 84,072 | 41 | -.13   | .051        | 24.20  | -.07   | -.10      | .029   | 37.01  |
| Autor(en) <sup>1a</sup> | ohne WPT    | 12,185 | 40 | -.07   | .079        | 37.81  | .04    | -.07      | .065   | 43.59  |
| CHC <sup>2</sup>        | WPT         | 5,253  | 27 | -.09   | .091        | 49.28  | .02    | -.07      | .075   | 47.82  |
| CHC <sup>2</sup>        | ohne WPT    | 87,172 | 49 | -.14   | .037        | 42.36  | -.10   | -.11      | .025   | 46.68  |
| CHC <sup>2a</sup>       | ohne WPT    | 18,674 | 48 | -.12   | .077        | 45.29  | -.02   | -.09      | .059   | 47.30  |
| <i>PSK</i>              |             |        |    |        |             |        |        |           |        |        |
| Autor(en) <sup>1</sup>  | FFM, NEO    | 11,173 | 42 | -.05   | .074        | 42.00  | .05    | -.05      | .068   | 44.85  |
| Autor(en) <sup>1</sup>  | NEO-PI-(R)  | 7,088  | 23 | -.02   | .068        | 39.91  | .06    | -.03      | .067   | 42.08  |
| Autor(en) <sup>1</sup>  | NEO-FFI     | 3,269  | 17 | -.11   | .064        | 60.32  | -.02   | -.09      | .062   | 57.15  |
| Autor(en) <sup>1</sup>  | FFM, no NEO | 4,264  | 17 | -.14   | .088        | 46.00  | -.03   | -.11      | .068   | 46.17  |
| Autor(en) <sup>1</sup>  | Eysenck     | 1,155  | 6  | -.11   | .000        | 100.00 | -.11   | -.08      | .000   | 100.00 |
| CHC <sup>2</sup>        | FFM, NEO    | 11,014 | 44 | -.07   | .066        | 54.57  | .01    | -.06      | .060   | 52.45  |
| CHC <sup>2</sup>        | NEO-PI-(R)  | 6,469  | 25 | -.04   | .060        | 56.72  | .03    | -.03      | .059   | 52.61  |
| CHC <sup>2</sup>        | NEO-FFI     | 4,545  | 19 | -.13   | .031        | 86.56  | -.09   | -.11      | .022   | 89.30  |
| CHC <sup>2</sup>        | FFM, no NEO | 4,735  | 17 | -.15   | .061        | 69.07  | -.07   | -.09      | .054   | 55.03  |
| CHC <sup>2</sup>        | Eysenck     | 4,413  | 12 | -.20   | .050        | 65.47  | -.13   | -.15      | .030   | 74.17  |

*Anmerkungen.* N = Stichprobengröße; k = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und –überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; % Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von Reeve et al. (2006, n = 71,887) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde und die Studie von Reeve et al. (2006, n = 71,887) exkludiert ist; kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; WPT = Wonderlic Personal Test (hierbei die Studien inkludiert, die den WPT zur Erfassung der allgemeinen Intelligenzausprägung verwendet haben; ohne WPT = die Studien inkludiert, die zur Erfassung der allgemeinen Intelligenz andere kogLT als den WPT verwendet haben; FFM, NEO = die Studien inkludiert, in denen Costa und McCrae-Inventare (NEO-PI, NEO-PI-R, NEO-FFI) eingesetzt wurden; NEO-PI-(R) = die Studien inkludiert, in denen der NEO-PI-R und der NEO-PI verwendet wurde; NEO-FFI = die Studien berücksichtigt, in denen der NEO-FFI eingesetzt wurde; FFM, no NEO = die Studien berücksichtigt, in denen FFM-Inventare, allerdings keine Costa und McCrae Inventare verwendet wurden; Eysenck = die Studien inkludiert, die Ne mit Eysenck-Inventaren erfasst haben; grau markiert sind die Zeilen, in denen mehr als 75% der Varianz erklärt wird.

## *Kapitel 5 – Ergebnisse*

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde, wird in den Studien, in denen Eysenck-Inventare eingesetzt wurden, mehr als 75% der Varianz erklärt (Tabelle 15, grau markiert). Dabei kann von einem negativen Zusammenhang von  $\rho = -.11$  für diese Subpopulation ausgegangen werden. Der Unterschied zwischen Eysenck- und Nicht-Eysenck-Inventaren führt zu erheblicher Varianz der Effektstärken. Offenbar sind die konzeptionellen Unterschiede zwischen dem FFM und Eysenck in Bezug auf N relevant für die Zusammenhänge mit g.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie erfolgte, wird in den Studien, in denen das NEO-FFI-Inventar eingesetzt wurde, mehr als 75% der Varianz erklärt. Die artefaktkorrigierte Effektgröße beträgt in dem Fall  $\rho = -.13$ . Es kann davon ausgegangen werden, dass die konzeptionellen Unterschiede zwischen den NEO-FFI-Inventaren und den Nicht-NEO-FFI-Inventaren in Bezug auf N relevant für die Zusammenhänge mit g sind.

### *Zusammenfassung zum Zusammenhang zwischen der Allgemeinen Intelligenz und des Neurotizismus:*

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der publizierten Studien beibehalten wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = -.13$ . Nach der Exklusion der Studie von Reeve et al. (2006) mit einer Stichprobengröße von  $n = 71,887$  verringert sich der Betrag der artefaktkorrigierten Effektgröße auf  $\rho = -.09$ . Dabei sollte die Richtung der artefaktkorrigierten Effektgröße nicht mehr generalisiert werden, da das Glaubwürdigkeitsintervall die Null mit einschließt. Unabhängig von der Exklusion bzw. Inklusion der Studie von Reeve et al. (2006) beträgt die erklärte Varianz weniger als 75% (Tabelle 10). Offenbar sind die korrigierten Effektgrößen ein mittlerer Effekt heterogener Subpopulationen.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durch den Autor der vorliegenden Arbeit erfolgte, beträgt die korrigierte Effektgröße  $\rho = -.14$ . Nach der Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) verringert sich der Betrag der korrigierten Effektgröße auf  $\rho = -.11$ . Unabhängig von der Exklusion bzw. Inklusion der Studie von Reeve et al. (2006) kann von einem negativen Effekt in der Population ausgegangen werden, da die Null nicht im Glaubwürdigkeitsintervall liegt. Darüber hinaus beträgt die erklärte Varianz auch in dieser Analyse-Variante deutlich weniger als 75% (Tabelle 10), weshalb ein mittlerer Effekt heterogener Subpopulationen nicht auszuschließen ist.

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der publizierten Studien beibehalten wurde, kann aufgrund des Schätzers mit  $R_0 < 1$  eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.07$  angenommen und ein Publikationsbias ausgeschlossen werden. Aufgrund der Höhe des zweiten Schätzers mit  $L_0 = 1.46$  wurden zwei Effektgrößen mit negativen Vorzeichen und den höchsten Beträgen exkludiert. Danach konnte eine Symmetrie hinsichtlich der Summe aller Ränge publizierter Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.07$  mit  $L_0 < 1$  angenommen und ein Publikationsbias ausgeschlossen werden. Nach der durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens, sollte tendenziell ein Publikationsbias nicht ausgeschlossen werden, da zu viele signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt sind. Um ein Publikationsbias ausschließen zu können, müssten 11 weitere nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, kann nach der „trim- and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) mit  $R_0 < 1$  und  $L_0 < 1$  eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.07$  angenommen und ein Publikationsbias ausgeschlossen werden.

Bezüglich der Moderatoranalysen, zeigte sich in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde, dass der Unterschied zwischen Eysenck- und Nicht-Eysenck-Inventaren zu einer erheblichen Varianz der Effektstärken führt. In den Studien, die Eysenck-Inventare zur Erfassung der Neurotizismusausprägung eingesetzt haben ( $k = 6$ ), konnten mehr als 75% der Varianz erklärt werden. In der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) durchgeführt wurde, konnte in den Studien, die den NEO-FFI zur Erfassung der Neurotizismusausprägung verwendet haben ( $k = 19$ ), mehr als 75% der Varianz erklärt werden.

## 5.2 Ergebnisse zur Fluiden Intelligenz und Neurotizismus

*Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien:*

In dieser Variante beträgt die inkludierte Studienanzahl  $k = 26$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = -.04$  ( $SD = .03$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 1,052$  (Range:  $n = 57$  bis  $20,175$ ;  $SD = 3,934.91$ ;  $N = 27,343$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. –überschätzung korrigierte

## Kapitel 5 – Ergebnisse

Effektgröße beträgt  $\rho = -.04$  ( $SD_\rho = .04$ ). Die Null liegt hierbei im Glaubwürdigkeitsintervall. Die erklärte Varianz beträgt 47.90% (Tabelle 15).

Durch die anschließende Exklusion der Studie von De Fruyt et al. (2006,  $n = 20,175$ ) verringerte sich der Betrag der unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = -.04$  ( $SD = .03$ ) auf  $\bar{r}_{bare-bones} = -.03$  ( $SD = .05$ ). Die mittlere Stichprobengröße beträgt hierbei  $\bar{n} = 286$  (Range:  $n = 57$  bis 2,658;  $SD = 517.96$ ;  $N = 7,168$ ). Der artefaktkorrigierte Effektgröße veränderte sich von  $\rho = -.04$  ( $SD_\rho = .04$ ) auf  $\rho = -.02$  ( $SD_\rho = .05$ ). Die Null liegt hier ebenfalls im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 52.46% (Tabelle 15). Nach Kemery et al. (1987) sollte bei einem derartigen Befund, unabhängig der Berücksichtigung der Studie von De Fruyt et al. (2006,  $n = 20,175$ ) ein mittlerer Effekt heterogener Subpopulationen angenommen werden, bei der die Höhe sowie die Richtung der Effektgröße nicht verallgemeinert sollte, da das Glaubwürdigkeitsintervall die Null mit einschließt und die erklärte Varianz kleiner als 75% ist. Dementsprechend wurden hier Moderatoranalysen durchgeführt (siehe unten).

**Tabelle 15**

Metaanalytische Ergebnisse zum Zusammenhang zwischen Gf und Ne

| Gf/Neurotizismus                     | $N$    | $k$ | $\rho$ | $SD_\rho$ | %Var. | 90% CV | $\bar{r}$ | $SD_r$ | %Var  |
|--------------------------------------|--------|-----|--------|-----------|-------|--------|-----------|--------|-------|
| kodiert nach Autor(en) <sup>1</sup>  | 27,343 | 26  | -.04   | .041      | 47.90 | .01    | -.04      | .027   | 56.55 |
| kodiert nach Autor(en) <sup>1a</sup> | 7,168  | 25  | -.02   | .053      | 52.46 | .05    | -.03      | .053   | 55.30 |
| kodiert nach CHC <sup>2</sup>        | 38,919 | 55  | -.05   | .053      | 43.38 | .02    | -.04      | .041   | 45.57 |
| kodiert nach CHC <sup>2a</sup>       | 18,744 | 54  | -.04   | .065      | 43.52 | .04    | -.04      | .060   | 44.76 |

*Anmerkungen.*  $N$  = Stichprobengröße;  $k$  = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und –überschätzung korrigierte Effektgröße;  $SD_\rho$  = Standardabweichung der artefaktkorrigierten Effektgröße; %Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analyse); <sup>1</sup> = Datensätze, in denen eine Zuordnung von Gf explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von Gf explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von De Fruyt et al. (2006,  $n = 20,175$ ) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde und die Studie von De Fruyt et al. (2006,  $n = 20,175$ ) exkludiert ist.

### Zuordnung der Fähigkeitskonstrukte gemäß CHC-Theorie:

In dieser Analyse-Variante beträgt die inkludierte Studienzahl  $k = 55$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = -.04$  ( $SD = .04$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 719$  (Range:  $n = 57$  bis 20,175;  $SD = 2,732.22$ ;  $N = 38,919$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. –überschätzung

## Kapitel 5 – Ergebnisse

korrigierte Effektgröße beträgt  $\rho = -.05$  ( $SD_\rho = .05$ ). Die Null liegt hierbei im Glaubwürdigkeitsintervall. Die erklärte Varianz beträgt 43.38% (Tabelle 15).

Durch die anschließende Exklusion der Studie von De Fruyt et al. (2006;  $n = 20,175$ ) verändert sich die unkorrigierte Effektgröße von  $\bar{r}_{bare-bones} = -.04$  nicht. Die mittlere Stichprobengröße beträgt hierbei  $\bar{n} = 571$  (Range:  $n = 57$  bis 2,982;  $SD = 570.75$ ;  $N = 18,744$ ). Der Betrag der artefaktkorrigierten Effektgröße verringert sich von  $\rho = -.05$  ( $SD_\rho = .05$ ) auf  $\rho = -.04$  ( $SD_\rho = .065$ ). Die Null liegt hier ebenfalls im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 43.52%. (Tabelle 15). Dementsprechend wurden hier Moderatoranalysen durchgeführt (siehe unten). Unabhängig von der Exklusion der Studie von De Fruyt et al. (2006) kann auch in dieser Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) erfolgte, angenommen werden, dass die artefaktkorrigierte Effektgröße ein mittlerer Effekt heterogener Subpopulationen ist und die Höhe sowie die Richtung der Effektgröße nicht verallgemeinert werden sollte.

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe der Schätzer mit  $R_0 < 1$  und  $L_0 < 1$  eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.01$  angenommen werden. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 25 Stichprobengrößen real publizierter Studien und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 16 aufgeführt.

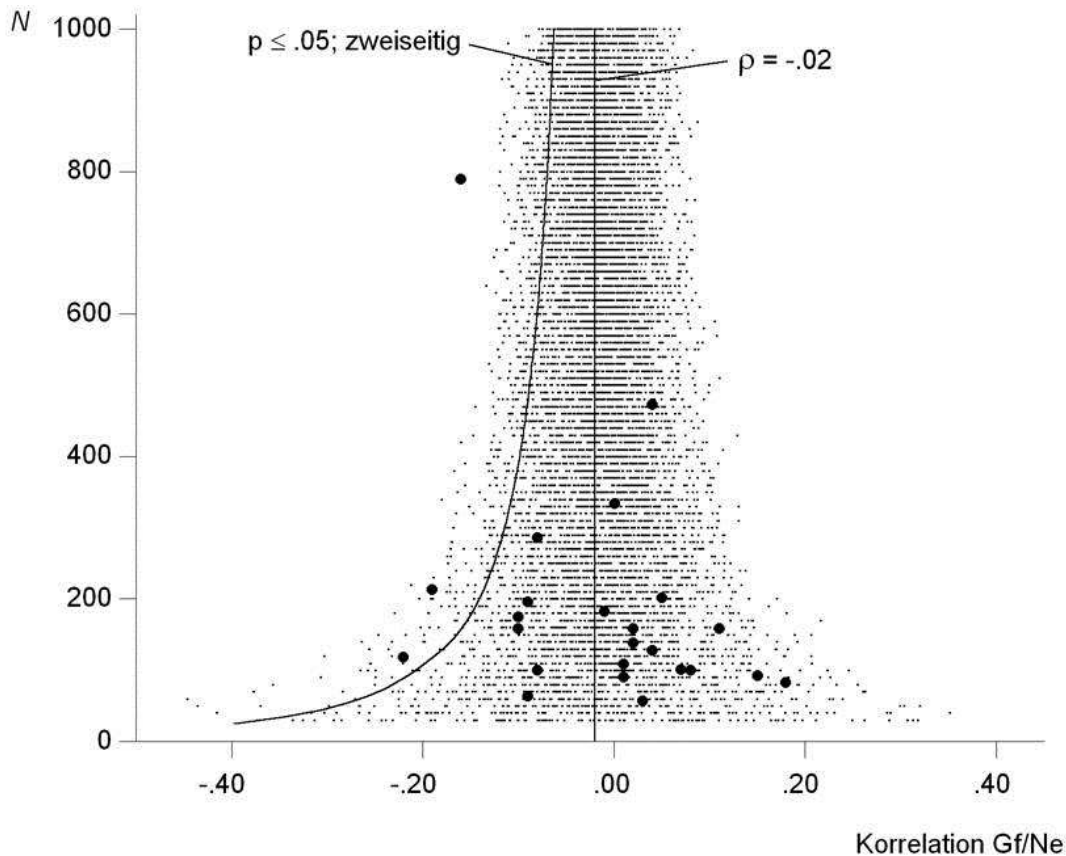
**Tabelle 16**

Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | $N$           | $k$ (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | $k$ (Simulation) <sup>3</sup> |
|-----------|---------------|------------------------------|-----------------------|-------------------------------|
| 1.        | 60 – 100      | 8                            | 32.00                 | 1,600                         |
| 2.        | 110 – 150     | 4                            | 16.00                 | 800                           |
| 3.        | 160 – 200     | 7                            | 28.00                 | 1,400                         |
| 4.        | 210 – 250     | 1                            | 4.00                  | 385                           |
| 5.        | 260 – 300     | 1                            | 4.00                  | 200                           |
| 6.        | 310 – 350     | 1                            | 4.00                  | 200                           |
| 7.        | 460 – 500     | 1                            | 4.00                  | 200                           |
| 8.        | 760 – 800     | 1                            | 4.00                  | 200                           |
| 9.        | 2,660 – 2,700 | 1                            | 4.00                  | 200                           |
| $\Sigma$  |               | <b>25</b>                    | <b>100.00</b>         | <b>5,002<sup>4</sup></b>      |

Anmerkungen.  $p = .02$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten  $k$  für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt  $k$  von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt  $k = 5,002$ .

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .05 berechnet sich aus einem Verhältnis von 233 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,769 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig). Da die simulierten Effektgrößen keinerlei Publikationsbias aufweisen, kann hierbei ein Verhältnis von 5% signifikanten ( $p \leq .05$ , zweiseitig) zu 95% nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) als Referenz zur Beurteilung eines Publikationsbias herangezogen werden. Der aus den tatsächlich publizierten Effektgrößen berechnete Index beträgt .12 (3 signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 22 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 16).



**Abbildung 16.**  $\rho = -.02$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 16) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien.

Um bei den tatsächlich publizierten Effektgrößen einen Index von .05 – ohne Publikationsbias – zu erreichen, müssten weitere 31 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einen Verhältnis von drei signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 53 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Dementsprechend kann ein Publikationsbias, bei dem zu viele signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sind, nicht ausgeschlossen werden.

Eine der publizierten Effektgrößen ( $r = -.16$ ;  $n = 789$ ; Beauducel et al., 2007) liegt im linken Bereich, außerhalb der simulierten Effektgrößen (Abbildung 16). In dieser Studie, in der die Auswirkungen des Zusammenhangs zwischen dem FFM und Gf sowie Gc mittels verschiedener Messmodelle untersucht wird, wird zur Erfassung der fluiden Intelligenzausprägung der Intelligenz-Struktur-Test 2000R (IST-2000R; Liepmann et al., 2008) und zur Erfassung der Persönlichkeitsdimension Ne der NEO-FFI in einer deutschen Version eingesetzt (Borkenau & Ostendorf, 1993). Das Alter der Probanden beträgt  $M = 18.0$  ( $SD = 1.38$ ) und besteht aus Schülern

unterschiedlicher Schulformen. Zur Erfassung von Gf mit dem IST-2000R werden insbesondere numerische und figurale Aufgabentypen berücksichtigt. Dabei erhalten die numerischen Aufgaben eine höhere Gewichtung als die figuralen Aufgaben. Denn neben der berichteten Gesamtkorrelation von  $r = -.16$  zwischen Gf und Ne, wird eine Korrelation zwischen Ne und der numerischen Intelligenz mit  $r = -.17$  sowie zwischen Ne und der figuralen Intelligenz mit  $r = -.07$  berichtet (Beauducel et al., 2007). Dementsprechend kann vermutet werden, dass sich insbesondere numerische Aufgabentypen leistungsmindernd auf Personen mit höheren Neurotizismuswerten auswirken können, was mittels Strukturgleichungsmodellen in Beauducel et al. (2007) gezeigt wird.

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durchgeführt wurde*

Nach der „trim and fill“ – Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann aufgrund der Höhe der Schätzer mit  $R_0 = 2$  und  $L_0 = 1.21$  eine asymmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.08$  angenommen werden. Für  $R_0$  wurden zwei „trim“-Schritte durchgeführt, bei denen insgesamt fünf negative Effektgrößen mit den höchsten Beträgen exkludiert wurden. Danach konnte mit  $R_0 < 1$  eine symmetrische Verteilung der Effektgrößen um die mittlere, neu bestimmte Effektgröße von  $\bar{r} = -.04$  angenommen werden. Für  $L_0$  wurden ebenfalls zwei „trim“-Schritte durchgeführt, bei denen insgesamt vier negative Effektgrößen mit den höchsten Beträgen exkludiert wurden. Danach konnte mit  $L_0 < 1$  eine Symmetrie hinsichtlich der Summe aller Ränge publizierter Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.04$  angenommen werden.

Darüber hinaus wurde in dieser Variante ebenfalls eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 54 Stichprobengrößen real publizierter Studien und die entsprechende Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 17 aufgeführt.



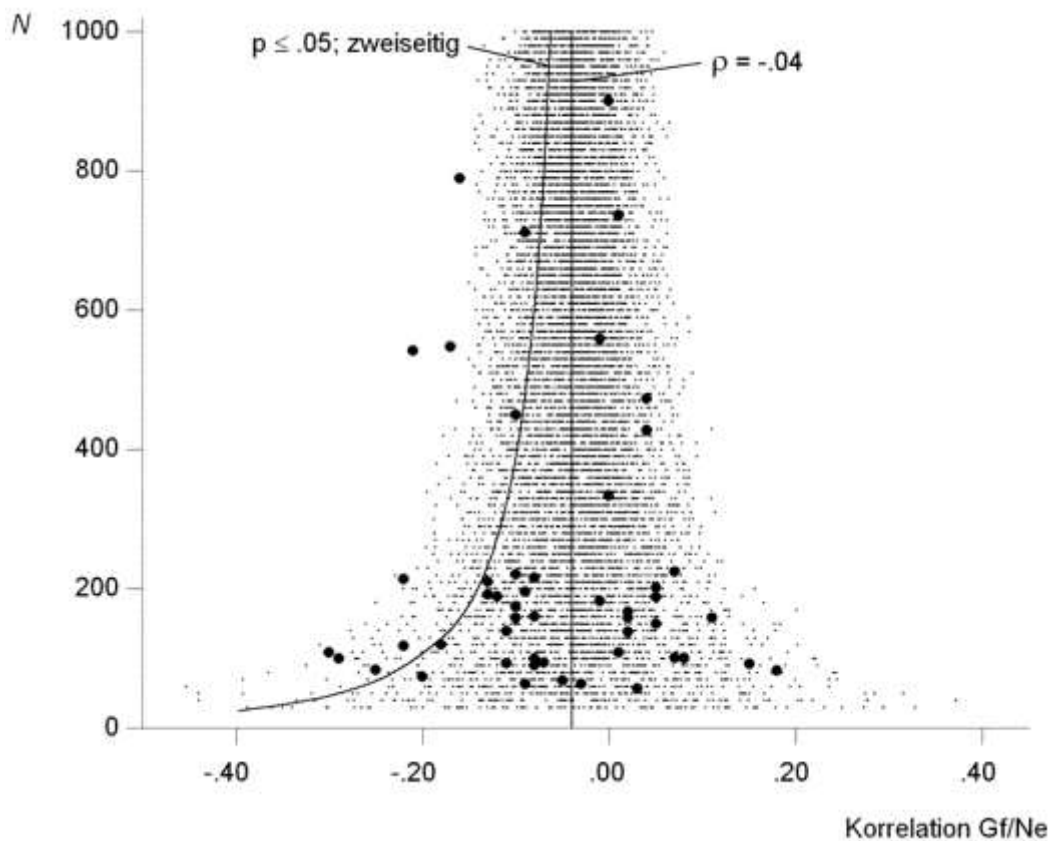
**Tabelle 17**

Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | $N$           | $k$ (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | $k$ (Simulation) <sup>3</sup> |
|-----------|---------------|------------------------------|-----------------------|-------------------------------|
| 1.        | 60 – 100      | 16                           | 29.63                 | 1,482                         |
| 2.        | 110 – 150     | 7                            | 12.96                 | 648                           |
| 3.        | 160 – 200     | 12                           | 22.22                 | 1,111                         |
| 4.        | 210 – 250     | 5                            | 9.26                  | 463                           |
| 5.        | 310 – 350     | 1                            | 1.85                  | 93                            |
| 6.        | 410 – 450     | 2                            | 3.70                  | 185                           |
| 7.        | 460 – 500     | 1                            | 1.85                  | 93                            |
| 8.        | 510 – 550     | 2                            | 3.70                  | 185                           |
| 9.        | 560 – 600     | 1                            | 1.85                  | 93                            |
| 10.       | 710 – 750     | 2                            | 3.70                  | 185                           |
| 11.       | 760 – 800     | 1                            | 1.85                  | 93                            |
| 12.       | 860 – 900     | 1                            | 1.85                  | 93                            |
| 13.       | 1,810 – 1,850 | 1                            | 1.85                  | 93                            |
| 14.       | 2,660 – 2,700 | 1                            | 1.85                  | 93                            |
| 15.       | 2,950 – 3,000 | 1                            | 1.85                  | 93                            |
| $\Sigma$  |               | <b>54</b>                    | <b>100.00</b>         | <b>5,019<sup>4</sup></b>      |

Anmerkungen.  $\rho = -.04$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten  $k$  für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt  $k$  von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt  $k = 5,019$ .

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .09 hat sich aus einem Verhältnis von 449 signifikanten ( $p \leq .05$ , zweiseitig) zu 4,570 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Im Vergleich dazu beträgt der Index der tatsächlich publizierten Effektgrößen .19 (10 signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 44 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 17).



**Abbildung 17.**  $\rho = -.04$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 17) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie (McGrew, 2009).

Um in den tatsächlich publizierten Effektgrößen einen Index von .09 – ohne Publikationsbias – zu erreichen, müssten weitere 52 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einen Verhältnis von 10 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 96 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Dementsprechend kann auch in dieser Analyse-Variante ein Publikationsbias, bei dem zu viele signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sind, nicht ausgeschlossen werden.

Drei der tatsächlich publizierten Effektgrößen liegen im linken Bereich, außerhalb der simulierten Effektgrößen (Abbildung 17; Tabelle 18). Auf mögliche Besonderheiten, weshalb die publizierte Effektgröße aus Beauducel et al. (2007) von den simulierten Effektgrößen abweicht, sind bereits unter Gf und Ne - Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien – aufgeführt (siehe oben). Interessanterweise werden in der Studie von Steinmayr, Dinger und Spinath (2010) ebenfalls wie in Beauducel et al. (2007) Schüler mit dem IST-2000R und dem

NEO-FFI untersucht. Hierbei korreliert die numerische Intelligenz mit Ne ebenfalls höher ( $r = -.19$ ) als die figurale Intelligenz ( $r = -.14$ ). Dementsprechend kann auch hier angenommen werden, dass sich insbesondere die numerischen Aufgabentypen leistungsmindernd auf Personen mit höheren Neurotizismuswerten auswirken. In der dritten Studie (Gow et al., 2005) in der eine Effektgröße ( $r = -.21$ ) berichtet wird, die außerhalb der simulierten Effektgrößen liegt (Abbildung 17), lassen sich keine eindeutigen Auffälligkeiten feststellen (Tabelle 18).

**Tabelle 18**

Deskriptive Angaben publizierter Studien, in denen die Effektgröße des Zusammenhangs zwischen Gf und Ne von den gezogenen Effektgrößen aus der adaptierten Simulation abweicht

| Studie   | kogLT     | PSK     | Land | Alter                          | Stichprobe                         | Thema  |
|--|-----------|---------|------|--------------------------------|------------------------------------|--|
| Beauducel, Liepmann, Felfe und Nettelstroth (2007) | IST-2000R | NEO-FFI | D    | $M = 18.00$<br>( $SD = 1.38$ ) | Schüler<br>( $n = 789$ )           | Auswirkungen verschiedener Messmodelle auf den Zusammenhang von Gf / Gc und Persönlichkeit |
| Gow, Whiteman, Pattie und Deary (2005)             | MHT       | IPIP    | GB   | $M = 79.00$<br>( $SD = 1.38$ ) | ältere Erwachsene<br>( $n = 542$ ) | Längsschnittstudie Zusammenhang Persönlichkeit und Intelligenz                             |
| Steinmayr, Dinger und Spinath (2010)               | IST-2000R | NEO-FFI | D    | $M = 16.98$<br>( $SD = 0.74$ ) | Schüler<br>( $n = 548$ )           | Familiäre Erziehung und schulischer Erfolg   |

Anmerkungen. kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; IST-2000R = Intelligenz-Struktur-Test 2000R; MHT = Moray House Test; IPIP = International Personality Item Pool (50 Items).

*Moderatoranalysen:*

Zum Zusammenhang zwischen Gf und Ne beträgt die Varianz des artefaktkorrigierten Effektes ( $\rho$ ) ausnahmslos weniger als 75% (Tabelle 15). Dementsprechend wurden Moderatoranalysen durchgeführt. Die kogLT berücksichtigend, wurden die Studien, in denen der Baddeley Reasoning Test (BRT; Baddeley, 1968) eingesetzt wurde, mit den Studien verglichen, die zur Erfassung der fluiden Intelligenzausprägung andere kogLT als den BRT verwendet haben (bspw. Culture Fair Intelligence Test, Cattell, 1973; Tabelle 19). Darüber hinaus wurden die Studien, in denen die Raven Tests (bspw. Raven et al., 2000) zur Erfassung der fluiden Intelligenzausprägung eingesetzt wurde, mit den Studien verglichen, in denen zur Messung der fluiden Intelligenzausprägung andere kogLT als die Raven Tests eingesetzt wurden (Tabelle 19). Des Weiteren wurden Moderatoranalysen hinsichtlich der eingesetzten Persönlichkeitsinventare durchgeführt. Dabei wurden folgende Persönlichkeitsinventare verglichen:

- FFM-Inventare (bspw. NEO-PI-(R), NEO-FFI, International Personality Item Pool) vs.
- NEO-Inventare (NEO-PI-(R) oder NEO-FFI) vs.
- NEO-PI-(R) vs.

Kapitel 5 – Ergebnisse

- NEO-FFI vs.
- Eysenck-Inventare.

In Tabelle 19 sind die metaanalytischen Kennwerte der entsprechenden Moderatoranalysen zum Zusammenhang zwischen Gf und Ne aufgeführt.

**Tabelle 19**  
Moderatoranalysen zum Zusammenhang zwischen Gf und Ne

| Gf/Ne                   | Moderator  | <i>N</i> | <i>k</i> | $\rho$ | $SD_\rho$ | % Var. | 90% CV | $\bar{r}$ | $SD_r$ | % Var  |
|-------------------------|------------|----------|----------|--------|-----------|--------|--------|-----------|--------|--------|
| <i>kogLT</i>            |            |          |          |        |           |        |        |           |        |        |
| Autor(en) <sup>1</sup>  | BRT        | 1,681    | 12       | .02    | .037      | 85.90  | -.02   | .03       | .030   | 88.66  |
| Autor(en) <sup>1</sup>  | ohne BRT   | 25,662   | 14       | -.05   | .035      | 41.83  | .00    | -.04      | .021   | 56.14  |
| Autor(en) <sup>1a</sup> | ohne BRT   | 5,487    | 13       | -.03   | .051      | 43.05  | .03    | -.05      | .046   | 52.41  |
| Autor(en) <sup>1</sup>  | Raven      | 21,394   | 6        | -.06   | .015      | 72.28  | -.04   | -.04      | .011   | 68.98  |
| Autor(en) <sup>1a</sup> | Raven      | 1,219    | 5        | -.05   | .063      | 60.82  | .03    | -.04      | .055   | 57.65  |
| Autor(en) <sup>1</sup>  | ohne Raven | 5,949    | 20       | -.02   | .051      | 51.78  | .05    | -.03      | .053   | 54.89  |
| CHC <sup>2</sup>        | BRT        | 1,681    | 12       | .02    | .037      | 85.90  | -.02   | .03       | .030   | 88.66  |
| CHC <sup>2</sup>        | ohne BRT   | 37,238   | 43       | -.05   | .051      | 40.84  | .01    | -.04      | .039   | 43.30  |
| CHC <sup>2a</sup>       | ohne BRT   | 17,063   | 42       | -.04   | .064      | 40.40  | .04    | -.05      | .058   | 42.39  |
| CHC <sup>2</sup>        | Raven      | 23,997   | 15       | -.07   | .037      | 48.78  | -.02   | -.04      | .027   | 46.04  |
| CHC <sup>2a</sup>       | Raven      | 3,822    | 14       | -.09   | .078      | 47.42  | .01    | -.07      | .065   | 46.69  |
| CHC <sup>2</sup>        | ohne Raven | 14,922   | 40       | -.03   | .058      | 45.53  | .04    | -.03      | .056   | 46.33  |
| <i>PSK</i>              |            |          |          |        |           |        |        |           |        |        |
| Autor(en) <sup>1</sup>  | FFM        | 27,286   | 25       | -.04   | .041      | 46.24  | .01    | -.04      | .028   | 54.66  |
| Autor(en) <sup>1a</sup> | FFM        | 7,111    | 24       | -.02   | .054      | 50.47  | .05    | -.03      | .055   | 53.30  |
| Autor(en) <sup>1</sup>  | NEO        | 26,787   | 23       | -.04   | .039      | 47.70  | .01    | -.04      | .025   | 57.27  |
| Autor(en) <sup>1a</sup> | NEO        | 6,612    | 22       | -.01   | .049      | 55.23  | .05    | -.02      | .050   | 57.31  |
| Autor(en) <sup>1</sup>  | NEO-PI-(R) | 25,476   | 17       | -.04   | .017      | 79.77  | -.02   | -.03      | .000   | 100.00 |
| Autor(en) <sup>1a</sup> | NEO-PI-(R) | 5,301    | 16       | -.01   | .000      | 100.00 | -.01   | -.01      | .000   | 100.00 |
| Autor(en) <sup>1</sup>  | NEO-FFI    | 1,311    | 6        | -.10   | .159      | 24.54  | .10    | -.08      | .110   | 27.20  |
| CHC <sup>2</sup>        | FFM        | 36,735   | 43       | -.04   | .050      | 41.82  | .02    | -.04      | .039   | 43.60  |
| CHC <sup>2a</sup>       | FFM        | 16,560   | 42       | -.03   | .061      | 42.33  | .05    | -.03      | .059   | 42.36  |
| CHC <sup>2</sup>        | NEO        | 35,236   | 39       | -.04   | .043      | 47.40  | .02    | -.03      | .03    | 49.67  |
| CHC <sup>2a</sup>       | NEO        | 15,088   | 38       | -.02   | .051      | 50.60  | .04    | -.02      | .050   | 50.12  |
| CHC <sup>2</sup>        | NEO-PI-(R) | 32,481   | 27       | -.03   | .018      | 79.92  | -.01   | -.03      | .011   | 87.12  |
| CHC <sup>2a</sup>       | NEO-PI-(R) | 12,306   | 26       | -.01   | .000      | 100.00 | -.01   | -.01      | .000   | 100.00 |
| CHC <sup>2</sup>        | NEO-FFI    | 2,755    | 12       | -.11   | .122      | 30.71  | .05    | -.09      | .100   | 31.64  |
| CHC <sup>2</sup>        | Eysenck    | 1,884    | 10       | -.08   | .045      | 81.52  | -.03   | -.07      | .029   | 86.23  |

Anmerkungen. *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und –überschätzung korrigierte Effektgröße;  $SD_\rho$  = Standardabweichung der artefaktkorrigierten Effektgröße; % Var. = erklärte Varianz; 90% CV = untere Grenze

## Kapitel 5 – Ergebnisse

des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, in denen eine Zuordnung von Gf explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von Gf explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von De Fruyt et al. (2006,  $n = 20,175$ ) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde und die Studie von De Fruyt et al. (2006,  $n = 20,175$ ) exkludiert ist; kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; BRT = Baddeley Reasoning Test (hierbei die Studien inkludiert, die den BRT zur Erfassung der fluiden Intelligenzausprägung verwendet haben; ohne BRT = die Studien inkludiert, die zur Erfassung der fluiden Intelligenz andere kogLT als den BRT verwendet haben; Raven = die Studien inkludiert, die den BRT zur Erfassung der fluiden Intelligenzausprägung verwendet haben; ohne Raven = die Studien inkludiert, die zur Erfassung der fluiden Intelligenz andere kogLT als den Raven Test verwendet haben; FFM = die Studien inkludiert, in denen „Big-Five“-Inventare eingesetzt wurden; NEO = die Studien inkludiert, in denen ausschließlich Costa und McCrae-Inventare (NEO-PI, NEO-PI-R, NEO-FFI) eingesetzt wurden; NEO-PI-(R) = die Studien inkludiert, in denen der NEO-PI-R und der NEO-PI verwendet wurde; NEO-FFI = die Studien berücksichtigt, in denen der NEO-FFI eingesetzt wurde; Eysenck = die Studien inkludiert, die Ne mit Eysenck-Inventaren erfasst haben; grau markiert sind die Zeilen, in denen mehr als 75% der Varianz erklärt wird.

Unabhängig von der Zuordnung der Fähigkeitskonstrukte (Autoren publizierter Studien vs. CHC) wird in den Studien, in denen der BRT verwendet wurde, mehr als 75% der Varianz erklärt (Tabelle 19, grau markiert). Die artefaktkorrigierte Effektgröße beträgt hierbei  $\rho = .02$ . Der Unterschied zwischen BRT- und Nicht-BRT-Inventaren führt zu erheblicher Varianz der Effektstärken. Offenbar sind die konzeptionellen Unterschiede zwischen dem BRT und nicht BRT in Bezug auf Gf relevant für die Zusammenhänge mit Ne.

Weiterhin wird in den Studien, in denen der NEO-PI-(R) zur Erfassung von Ne verwendet wurde, mehr als 75% der Varianz erklärt (Tabelle 19, grau markiert). Die artefaktkorrigierte Effektgröße beträgt für diese Subpopulation  $\rho = -.03$  bzw.  $\rho = -.01$ . Hierbei ist anzunehmen, dass die konzeptionellen Unterschiede zwischen dem NEO-PI-(R) und den Nicht-NEO-PI-(R)-Inventaren in Bezug auf Ne mit Gf relevant sind.

In den Studien, in denen ein Eysenck-Inventar zur Erfassung von Ne verwendet wurde, kann ebenfalls mehr als 75% der Varianz erklärt werden (Tabelle 19, grau markiert). Allerdings gilt dies ausschließlich für die Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde. Offenbar sind die konzeptionellen Unterschiede von Ne, die mit Eysenck-Inventaren in dieser Analyse-Variante erfasst werden, relevant in Bezug auf die Zusammenhänge mit Gf.

### *Zusammenfassung zum Zusammenhang zwischen der Fluiden Intelligenz und Neurotizismus:*

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien vorgenommen wurde, beträgt die korrigierte Effektgröße  $\rho = -.04$ . Nach der

Exklusion der Studie von De Fruyt et al. (2006) mit einer Stichprobengröße von  $n = 20,175$  verringert sich der Betrag der korrigierten Effektgröße auf  $\rho = -.02$ . Unabhängig von der Exklusion bzw. Inklusion der Studie von De Fruyt et al. (2006) beträgt die erklärte Varianz weniger als 75% und die Null liegt im Glaubwürdigkeitsintervall. Offenbar sind die korrigierten Effektgrößen ein mittlerer Effekt heterogener Subpopulationen und die Höhe sowie die Richtung der Effektgröße sollte nicht verallgemeinert werden.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durch den Autor der vorliegenden Arbeit vorgenommen wurde, beträgt die korrigierte Effektgröße  $\rho = -.05$ . Nach der Exklusion der Studie von De Fruyt et al. (2006,  $n = 20,175$ ) verringert sich der Betrag der korrigierten Effektgröße auf  $\rho = -.04$ . Auch in dieser Analyse-Variante beträgt die erklärte Varianz deutlich weniger als 75% und die Null liegt im Glaubwürdigkeitsintervall (Tabelle 15). Dementsprechend kann unabhängig von der Exklusion bzw. Inklusion der Studie von De Fruyt et al. (2006) ein mittlerer Effekt heterogener Subpopulationen angenommen werden, bei der die Höhe sowie die Richtung der Effektgröße nicht verallgemeinert werden sollte.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien berücksichtigt wurde, kann nach der „trim- and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.01$  angenommen und ein Publikationsbias ausgeschlossen werden. Nach der durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens, kann hingegen ein Publikationsbias nicht ausgeschlossen werden, da zu viele signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt sind. Um ein Publikationsbias ausschließen zu können, müssten 31 weitere nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) publiziert worden sein.

In der Analyse-Variante, eine der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, kann nach der „trim- and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) ein Publikationsbias nicht ausgeschlossen werden. Nach der Exklusion von fünf bzw. vier Effektgrößen ( $p \leq .05$ , zweiseitig) mit negativen Vorzeichen und den größten Beträgen, kann für die Schätzer mit  $R_0 < 1$  und  $L_0 < 1$  eine symmetrische Verteilung der publizierten Effektgrößen angenommen und ein Publikationsbias ausgeschlossen werden. Nach der durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens, ist tendenziell ein Publikationsbias nicht auszuschließen, da zu viele signifikante Effektgrößen ( $p \leq$

.05, zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt sind. Um ein Publikationsbias ausschließen zu können, müssten 52 weitere nichtsignifikante Effektgrößen ( $p \leq .05$ , zweiseitig) publiziert sein.

Bezüglich der Moderatoranalysen konnte unabhängig der Zuordnung der Fähigkeitskonstrukte (Autoren vs. CHC) in den Studien, in denen der BRT oder der NEO-PI-R eingesetzt wurde, mehr als 75% der Varianz erklärt werden. Offensichtlich sind die konzeptionellen Unterschiede zwischen BRT- und Nicht-BRT-Inventaren sowie NEO-PI-R- und Nicht-NEO-PI-R-Inventaren relevant in Bezug auf den Zusammenhang zwischen Gf und Ne. In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, kann darüber hinaus durch den Moderator Eysenck-Inventar mehr als 75% der Varianz erklärt werden.

### 5.3 Ergebnisse zur Kristallisierten Intelligenz und Neurotizismus

*Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien:*

In der Variante beträgt die inkludierte Studienzahl  $k = 10$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = -.15$  ( $SD = .02$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 7,593$  (Range:  $n = 57$  bis 71,887;  $SD = 22,596.74$ ;  $N = 75,933$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. –überschätzung korrigierte Effektgröße beträgt  $\rho = -.20$  ( $SD_{\rho} = .03$ ). Die Null liegt hierbei nicht im Glaubwürdigkeitsintervall. Die erklärte Varianz beträgt 18.89% (Tabelle 20).

Nach der anschließenden Exklusion der Studie von Reeve et al. (2006;  $n = 71,887$ ) verringert sich der Betrag der unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = -.15$  ( $SD = .02$ ) auf  $-.06$  ( $SD = .00$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 450$  (Range:  $n = 57$  bis 1,821;  $SD = 563.15$ ;  $N = 4,046$ ). Der Betrag der artefaktkorrigierten Effektgröße verringert sich hierbei von  $\rho = -.20$  ( $SD_{\rho} = .03$ ) auf  $\rho = -.07$  ( $SD_{\rho} = .00$ ). Das Glaubwürdigkeitsintervall schließt die Null nicht ein und die erklärte Varianz beträgt 100% (Tabelle 20).

Generell kann von einem negativen Zusammenhang zwischen der kristallisierten Intelligenz und Neurotizismus ausgegangen werden, da die Null nicht im Glaubwürdigkeitsintervall liegt. Nach der Exklusion der Studie von Reeve et al. (2006) kann darüber hinaus die Höhe der artefaktkorrigierten Effektgröße von  $\rho = -.07$  generalisiert werden, da mehr als 75% der Varianz erklärt werden. Offensichtlich wird hierbei die Höhe der artefaktkorrigierten Effektgröße von einer

einzelnen Studie (Reeve et al., 2006) mit einer sehr großen Stichprobengröße von  $n = 71,887$  moderiert.

**Tabelle 20**

Metaanalytische Ergebnisse zum Zusammenhang zwischen Gc und Ne

| Gc/Neurotizismus                     | $N$    | $k$ | $\rho$ | $SD_{\rho}$ | %Var.  | 90% CV | $\bar{r}$ | $SD_r$ | %Var   |
|--------------------------------------|--------|-----|--------|-------------|--------|--------|-----------|--------|--------|
| kodiert nach Autor(en) <sup>1</sup>  | 75,933 | 10  | -.20   | .032        | 18.89  | -.16   | -.15      | .020   | 23.58  |
| kodiert nach Autor(en) <sup>1a</sup> | 4,046  | 9   | -.07   | .000        | 100.00 | -.07   | -.06      | .000   | 100.00 |
| kodiert nach CHC <sup>2</sup>        | 78,829 | 30  | -.19   | .047        | 23.56  | -.13   | -.14      | .026   | 35.40  |
| kodiert nach CHC <sup>2a</sup>       | 6,942  | 29  | -.07   | .042        | 74.93  | -.02   | -.07      | .035   | 76.91  |

*Anmerkungen.*  $N$  = Stichprobengröße;  $k$  = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und -überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; %Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analyse); <sup>1</sup> = Datensätze, in denen eine Zuordnung von Gc explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von Gc explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von Reeve et al. (2006,  $n = 71,887$ ) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde und die Studie von Reeve et al. (2006,  $n = 71,887$ ) exkludiert ist.

*Zuordnung der Fähigkeitskonstrukte gemäß CHC-Theorie:*

In dieser Analyse-Variante beträgt die inkludierte Studienzahl  $k = 30$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = -.14$  ( $SD = .03$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 2,628$  (Range:  $n = 57$  bis  $n = 71,887$ ;  $SD = 13,082.27$ ;  $N = 78,829$ ;). Die artefaktkorrigierte Effektgröße beträgt  $\rho = -.19$  ( $SD_{\rho} = .05$ ). Das Glaubwürdigkeitsintervall schließt die Null nicht ein und die erklärte Varianz beträgt 23.56% (Tabelle 20).

Durch die anschließende Exklusion der Studie von Reeve et al. (2006;  $n = 71,887$ ), verringert sich der Betrag der unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = -.14$  ( $SD = .03$ ) auf  $-.07$  ( $SD = .04$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 239$  (Range:  $n = 57$  bis  $n = 789$ ;  $SD = 185.06$ ;  $N = 6,942$ ;). Der Betrag der artefaktkorrigierten Effektgröße verringert sich hierbei von  $\rho = -.19$  ( $SD_{\rho} = .05$ ) auf  $-.07$  ( $SD_{\rho} = .04$ ). Die Null liegt hierbei nicht im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 74.93% (Tabelle 20).

Auch in dieser Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der der CHC-Theorie (McGrew, 2009) erfolgte, kann angenommen werden, dass die Höhe der artefaktkorrigierten Effektgröße durch eine einzelne Studie moderiert wird. Durch die Exklusion



der Studie von Reeve et al. (2006;  $n = 71,887$ ) kann die Höhe der artefaktkorrigierte Effektgröße von  $\rho = -.07$  mit einer inkludierten Stichprobenanzahl von immerhin  $k = 29$  über verschiedene Situationen hinweg generalisiert werden. Da nach der Exklusion der Studie von Reeve et al. (2006) die erklärte Varianz in beiden Varianten (Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien vs. CHC-Theorie) mindestens 74.93% beträgt, wurden keine Moderatoranalysen zum Zusammenhang zwischen Gc und Ne gerechnet.

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe der Schätzer mit  $R_0 < 1$  und  $L_0 < 1$  von einer symmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.05$  ausgegangen werden. Demnach ist ein Publikationsbias auszuschließen. Weiterhin wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der neun Stichprobengrößen real publizierter Studien und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 21 aufgeführt.

**Tabelle 21**

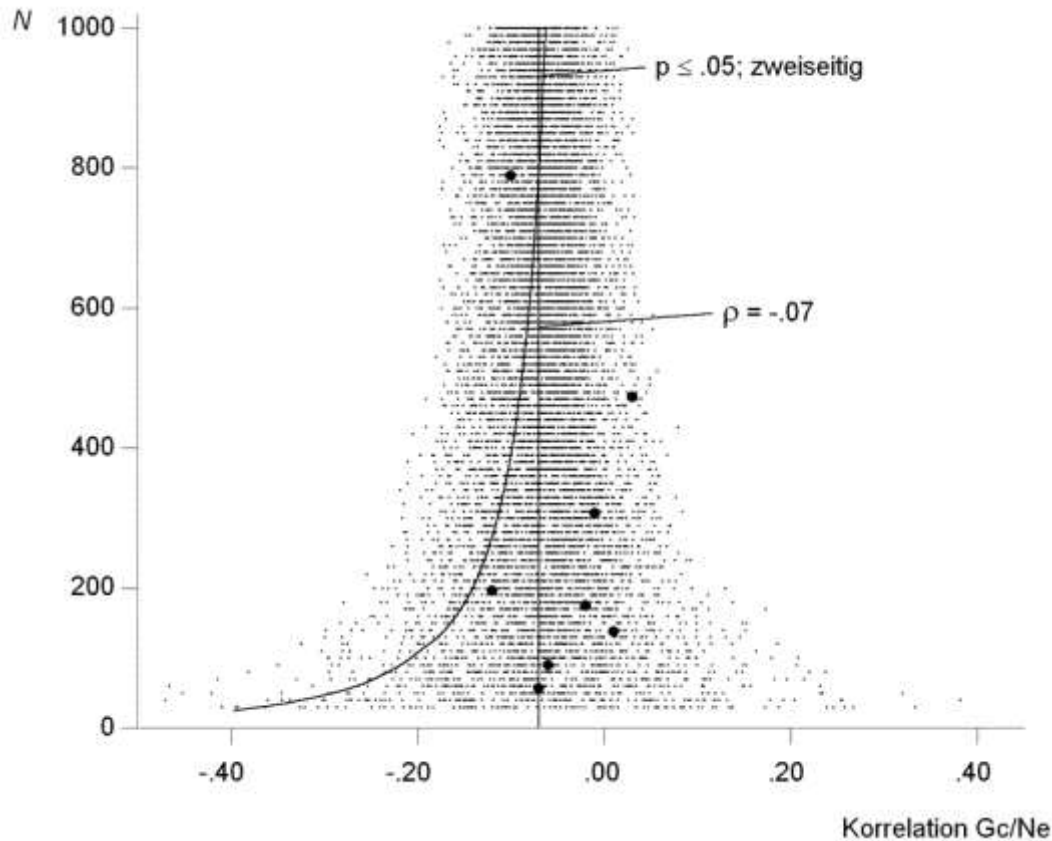
Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | $N$           | $k$ (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | $k$ (Simulation) <sup>3</sup> |
|-----------|---------------|------------------------------|-----------------------|-------------------------------|
| 1.        | 60 – 100      | 2                            | 22.22                 | 1,111                         |
| 2.        | 110 – 150     | 1                            | 11.11                 | 556                           |
| 3.        | 160 – 200     | 2                            | 22.22                 | 1,111                         |
| 4.        | 310 – 350     | 1                            | 11.11                 | 556                           |
| 5.        | 460 – 500     | 1                            | 11.11                 | 556                           |
| 6.        | 760 – 800     | 1                            | 11.11                 | 556                           |
| 7.        | 1,810 – 1,850 | 1                            | 11.11                 | 556                           |
| $\Sigma$  |               | <b>9</b>                     | <b>100.00</b>         | <b>5,002<sup>4</sup></b>      |

Anmerkungen.  $\rho = -.07$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten  $k$  für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt  $k$  von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt  $k = 5,002$ .

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .26 hat sich aus einem Verhältnis von 1,285 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 3,717 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Da die simulierten Effektgrößen keinerlei Publikationsbias aufweisen, kann hierbei ein Verhältnis von 26% signifikanten ( $p \leq .05$ , zweiseitig) zu 74% nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) als Referenz zur

Beurteilung eines Publikationsbias herangezogen werden. Der aus den tatsächlich publizierten Effektgrößen berechnete Index beträgt .22 (zwei signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; sieben nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 18).



**Abbildung 18.**  $\rho = -.07$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 21) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien.

Um bei den tatsächlich publizierten Effektgrößen einen Index von ungefähr .26 – ohne Publikationsbias – zu erreichen, müsste eine weitere signifikante Effektgröße ( $p \leq .05$ , zweiseitig) in einem Verhältnis von drei signifikanten ( $p \leq .05$ , zweiseitig) zu sieben nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Allerdings würde der Index durch eine zusätzliche signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) bereits .30 betragen. Dementsprechend kann hier ein Publikationsbias ausgeschlossen werden. Überdies ist anhand der Abbildung 18 sichtbar, dass die tatsächlich publizierten Effektgrößen mit  $n \leq 1,000$  innerhalb der simulierten Effektgrößen liegen.

Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durchgeführt wurde:

Nach der „trim and fill“ – Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann ein Publikationsbias nicht ausgeschlossen werden. Der Schätzer  $R_0$  zeigt eine asymmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.06$  an. Durch die Exklusion der drei Effektgrößen mit positiven Vorzeichen und den höchsten Beträgen kann eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere, neu berechnete Effektgröße von  $\bar{r} = -.08$  mit  $R_0 < 1$  angenommen werden. Für den zweiten Schätzers  $L_0$  waren zwei Effektgrößen mit negativen Vorzeichen und den höchsten Beträgen zu exkludieren, damit eine Symmetrie hinsichtlich der Summer aller Ränge publizierter Effektgrößen um die mittlere, neu berechnete Effektgröße von  $\bar{r} = -.05$  mit  $L_0 < 1$  angezeigt wird. Zusätzlich wurde eine Simulation gemäß des im Kapitel 4.6.7. beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 29 Stichprobengrößen real publizierter Studien und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 22 aufgeführt.

**Tabelle 22**

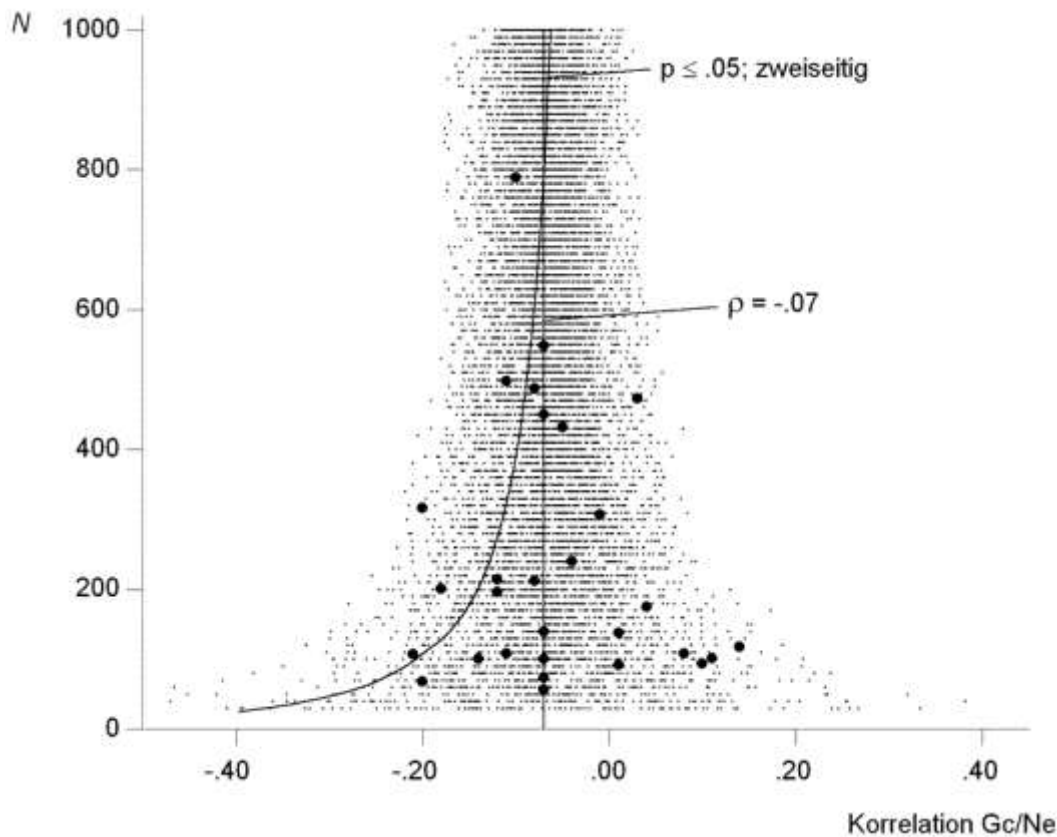
Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | $N$       | $k$ (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | $k$ (Simulation) <sup>3</sup> |
|-----------|-----------|------------------------------|-----------------------|-------------------------------|
| 1.        | 60 – 100  | 8                            | 27.59                 | 1,380                         |
| 2.        | 110 – 150 | 6                            | 20.69                 | 1,035                         |
| 3.        | 160 – 200 | 3                            | 10.34                 | 517                           |
| 4.        | 210 – 250 | 3                            | 10.34                 | 517                           |
| 5.        | 310 – 350 | 2                            | 6.90                  | 345                           |
| 6.        | 410 – 450 | 2                            | 6.90                  | 345                           |
| 7.        | 460 – 500 | 3                            | 10.34                 | 517                           |
| 8.        | 510 – 550 | 1                            | 3.45                  | 173                           |
| 9.        | 760 – 800 | 1                            | 3.45                  | 173                           |
| $\Sigma$  |           | <b>29</b>                    | <b>100.00</b>         | <b>5,002<sup>4</sup></b>      |

Anmerkungen.  $p = -.07$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten  $k$  für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt  $k$  von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt  $k = 5,002$ .

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .15 hat sich aus einem Verhältnis von 766 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) und 4,236 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Da die simulierten Effektgrößen keinerlei Publikationsbias aufweisen, kann hierbei ein Verhältnis von 15% signifikanten ( $p \leq .05$ , zweiseitig) zu 85% nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) als Referenz zur Beurteilung eines Publikationsbias herangezogen werden. Der aus den tatsächlich publizierten

Effektgrößen berechnete Index beträgt .17 (fünf signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 24 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 19).



**Abbildung 19.**  $\rho = -.07$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 22) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie (McGrew, 2009).

Um bei den tatsächlich publizierten Effektgrößen einen Index von .15 – ohne Publikationsbias – zu erreichen, müssten weitere fünf nichtsignifikante Effektgrößen ( $p \leq .05$ , zweiseitig) in einem Verhältnis von fünf signifikanten ( $p \leq .05$ , zweiseitig) zu 29 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Dementsprechend kann tendenziell ein Publikationsbias nicht ausgeschlossen werden. Offenbar sind mehr signifikante Effektgrößen mit negativen Vorzeichen ( $p \leq .05$ , zweiseitig) im Vergleich zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde. Darüber hinaus ist liegt keine der tatsächlich publizierten Effektgrößen mit  $n < 1,000$  außerhalb der simulierten Effektgrößen (Abbildung 19).

## Kapitel 5 – Ergebnisse

### *Zusammenfassung zum Zusammenhang zwischen der Kristallisierten Intelligenz und des Neurotizismus:*

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien beibehalten wurde, beträgt die korrigierte Effektgröße  $\rho = -.20$ . Die erklärte Varianz beträgt weniger als 20% (Tabelle 20). Offensichtlich ist die korrigierte Effektgröße ein mittlerer Effekt heterogener Subpopulationen. Nach der Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) verringert sich der Betrag der korrigierten Effektgröße auf  $\rho = -.07$ , die Null liegt nicht im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 100%. Dementsprechend kann bei einem derartigen Befund die Richtung und Höhe der artefaktkorrigierte Effektgröße generalisiert werden (vgl. Kemery et al., 1987).

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, beträgt die korrigierte Effektgröße  $\rho = -.19$ . Nach der Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) verringert sich der Betrag der artefaktkorrigierten Effektgröße auf  $\rho = -.07$ . Auch in dieser Analyse-Variante kann durch die Exklusion der Studie von Reeve et al. (2006) die Richtung und Höhe der artefaktkorrigierten Effektgröße von  $\rho = -.07$  generalisiert werden. In beiden Analyse-Varianten (Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien vs. CHC-Theorie) wird die Höhe der artefaktkorrigierte Effektgröße von einer einzelnen Studie mit einer sehr großen Stichprobe moderiert.

Ein Publikationsbias kann nach der „trim- and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien durchgeführt wurde, ausgeschlossen werden. Nach der durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens kann ebenfalls ein Publikationsbias ausgeschlossen werden.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, kann nach der „trim- and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) ist von einer asymmetrischen Verteilung der publizierten Effektgrößen auszugehen und ein Publikationsbias sollte dementsprechend nicht ausgeschlossen werden. Für den Schätzer  $R_0$  waren drei positive Effektgrößen mit den höchsten Beträgen zu exkludieren, um eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße zu erhalten. Für den zweiten Schätzer  $L_0$  waren hingegen zwei negative Effektgrößen mit den höchsten Beträgen zu exkludieren, um eine Symmetrie hinsichtlich der Summe aller Ränge publizierter

Effektgrößen um die mittlere Effektgröße zu erhalten. Um ein Publikationsbias mittels Simulation ausschließen zu können, müssten weitere fünf nichtsignifikante Effektgrößen ( $p \leq .05$ , zweiseitig) publiziert sein.

#### 5.4 Ergebnisse zur Allgemeinen Intelligenz und Extraversion

*Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien:*

In der Analyse-Variante beträgt die inkludierte Studienzahl  $k = 71$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .02$  ( $SD = .05$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 1,336$  (Range:  $n = 59$  bis  $71,887$ ;  $SD = 8,512.61$ ;  $N = 98,294$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. -überschätzung korrigierte Effektgröße beträgt  $\rho = .02$  ( $SD_{\rho} = .06$ ). Die Null liegt hierbei im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 25.04% (Tabelle 23).

Durch die anschließende Exklusion der Studie von Reeve et al. (2006) mit einer Stichprobengröße von  $n = 71,887$  verringert sich die unkorrigierte Effektgröße von  $\bar{r}_{bare-bones} = .02$  ( $SD = .05$ ) auf  $-.01$  ( $SD = .09$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 328$  (Range:  $n = 59$  bis  $4,547$ ;  $SD = 589.86$ ;  $N = 23,018$ ). Die artefaktkorrigierte Effektgröße verringert sich hierbei von  $\rho = .02$  ( $SD_{\rho} = .06$ ) auf  $\rho = -.02$  ( $SD_{\rho} = .10$ ). Das Glaubwürdigkeitsintervall schließt die Null ein und die erklärte Varianz beträgt 28.75% (Tabelle 23).

Unabhängig von der Exklusion bzw. Inklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) ist die erklärte Varianz deutlich kleiner als 75%. Dementsprechend wurden Moderatoranalysen durchgeführt (siehe unten).

**Tabelle 23**

Metaanalytische Ergebnisse zum Zusammenhang zwischen g und E

| g/Extraversion                       | <i>N</i> | <i>k</i> | $\rho$ | $SD_\rho$ | %Var. | 90% CV | $\bar{r}$ | $SD_r$ | %Var  |
|--------------------------------------|----------|----------|--------|-----------|-------|--------|-----------|--------|-------|
| kodiert nach Autor(en) <sup>1</sup>  | 98,294   | 71       | .02    | .062      | 25.04 | -.06   | .02       | .046   | 26.42 |
| kodiert nach Autor(en) <sup>1a</sup> | 23,018   | 70       | -.02   | .103      | 28.75 | .11    | -.01      | .087   | 28.64 |
| kodiert nach CHC <sup>2</sup>        | 98,717   | 86       | .02    | .064      | 28.13 | -.06   | .02       | .047   | 28.81 |
| kodiert nach CHC <sup>2a</sup>       | 26,830   | 85       | -.01   | .107      | 31.44 | .12    | -.01      | .083   | 31.68 |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und -überschätzung korrigierte Effektgröße;  $SD_\rho$  = Standardabweichung der artefaktkorrigierten Effektgröße; %Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analyse); <sup>1</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde; <sup>2a</sup> = Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde und die Studie von Reeve et al. (2006) exkludiert ist.

*Zuordnung der Fähigkeitskonstrukte gemäß CHC-Theorie:*

In dieser Variante beträgt die inkludierte Studienanzahl *k* = 86 bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .02$  ( $SD = .05$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 1,148$  (Range: *n* = 57 bis 71,887;  $SD = 7,736.61$ ; *N* = 98,294). Die artefaktkorrigierte Effektgröße beträgt  $\rho = .02$  ( $SD_\rho = .06$ ). Das Glaubwürdigkeitsintervall schließt die Null ein und die erklärte Varianz beträgt 28.13% (Tabelle 23).

Durch die anschließende Exklusion der Studie von Reeve et al. (2006; *n* = 71,887) verringert sich die unkorrigierte Effektgröße von  $\bar{r}_{bare-bones} = .02$  ( $SD = .05$ ) auf  $-.01$  ( $SD = .08$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 316$  (Range: *n* = 57 bis *n* = 4,547;  $SD = 543.20$ ; *N* = 26,830). Die artefaktkorrigierte Effektgröße verringert sich hierbei von  $\rho = .02$  ( $SD_\rho = .06$ ) auf  $-.01$  ( $SD_\rho = .11$ ). Das Glaubwürdigkeitsintervall schließt die Null auch nach der Exklusion von Reeve et al. (2006) mit ein und die erklärte Varianz beträgt 31.44% (Tabelle 23). Auch in dieser Analyse-Variante beträgt die erklärte Varianz, unabhängig von der Exklusion bzw. Inklusion Studie von Reeve et al. (2006, *n* = 71,887), deutlich weniger als 75%. Dementsprechend wurden auch in dieser Variante Moderatoranalysen durchgeführt (siehe unten).

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe des Schätzers  $R_0 < 1$  eine weitestgehend symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .04$  angenommen werden.

Demnach wäre kein Publikationsbias anzunehmen. Der zweite Schätzer  $L_0$  zeigt bei den Effektgrößen mit positiven Vorzeichen kein Publikationsbias an. Allerdings kann bei den publizierten Effektgrößen, die kleiner als die mittlere Effektgröße von  $\bar{r} = .04$  sind, eine Asymmetrie mit  $L_0 = 12.91$  angenommen werden. Demnach sind 13 Effektgrößen ( $L_0^+ = 13$ ) mit negativen Vorzeichen und den höchsten Beträgen zu exkludieren. Nach dem ersten „trim“-Schritt beträgt der Schätzer  $L_0 = 4.35$ . Demzufolge wurden in einer weiteren Iteration fünf Effektgrößen ( $L_0^+ = 5$ ) mit negativen Vorzeichen und den höchsten Beträgen exkludiert. Nach zwei „trim“-Schritten, in denen insgesamt 18 Effektgrößen exkludiert wurden, kann eine symmetrische Verteilung der Effektgrößen hinsichtlich der Summe aller Ränge publizierter Effektgrößen (siehe Anhang A4) um die mittlere, neu bestimmte Effektgröße von  $\bar{r} = .05$  mit  $L_0 < 1$  angenommen werden.

Neben den unterschiedlichen Ergebnissen hinsichtlich der Beurteilung eines Publikationsbias nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b), wurde zusätzlich eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 70 Stichprobengrößen real publizierter Studien und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 24 aufgeführt.

**Tabelle 24**

Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

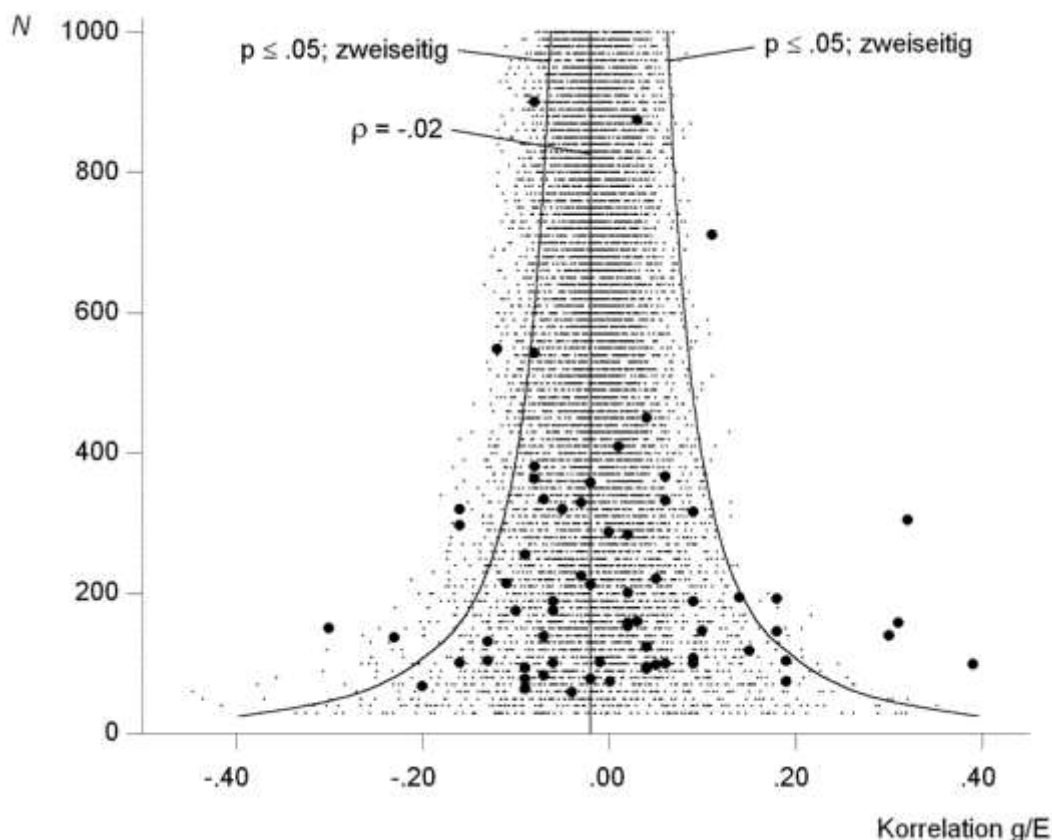
| Intervall | $N$           | $k$ (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | $k$ (Simulation) <sup>3</sup> |
|-----------|---------------|------------------------------|-----------------------|-------------------------------|
| 1.        | 60 – 100      | 20                           | 28.57                 | 1,429                         |
| 2.        | 110 – 150     | 11                           | 15.71                 | 786                           |
| 3.        | 160 – 200     | 10                           | 14.29                 | 715                           |
| 4.        | 210 – 250     | 4                            | 5.71                  | 286                           |
| 5.        | 260 – 300     | 4                            | 5.71                  | 286                           |
| 6.        | 310 – 350     | 7                            | 10.00                 | 493                           |
| 7.        | 360 – 400     | 4                            | 5.71                  | 286                           |
| 8.        | 410 – 450     | 2                            | 2.86                  | 143                           |
| 9.        | 510 – 550     | 2                            | 2.86                  | 143                           |
| 10.       | 710 – 750     | 1                            | 1.43                  | 72                            |
| 11.       | 860 – 900     | 2                            | 2.86                  | 143                           |
| 12.       | 1,310 – 1,350 | 1                            | 1.43                  | 72                            |
| 13.       | 1,860 – 1,900 | 1                            | 1.43                  | 72                            |
| 14.       | 4,510 – 4,550 | 1                            | 1.43                  | 72                            |
| $\Sigma$  |               | <b>70</b>                    | <b>100.00</b>         | <b>4,998<sup>4</sup></b>      |

Anmerkungen.  $\rho = -.02$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten  $k$  für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt  $k$  von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt  $k = 4,998$ .



Unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich hat sich aus der adaptierten Simulation ein Index von .05 berechnet (231 signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 4,767 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig). Hinsichtlich der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich hat sich aus der Simulation ein Index von .01 berechnet (51 signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 4,947 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 20).

Der aus den tatsächlich publizierten Effektgrößen berechnete Index unter Berücksichtigung der Signifikanzgrenze im negativen Wertebereich beträgt .11 (acht signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 62 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig). Der aus den tatsächlich publizierten Effektgrößen berechnete Index unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich beträgt ebenfalls .11 (acht signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 62 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 20).



**Abbildung 20.**  $\rho = -.02$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 24) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktionen im positiven und negativen Wertebereich mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien.

Um bei den tatsächlich publizierten Effektgrößen einen Index von .05 – ohne Publikationsbias – zu erreichen, müssten weitere 77 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von acht signifikanten ( $p \leq .05$ , zweiseitig) zu 140 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich publiziert sein. Unter Berücksichtigung der Signifikanzgrenze im positiven Wertebereich ( $p \leq .05$ , zweiseitig) müssten weitere 501 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von acht signifikanten ( $p \leq .05$ , zweiseitig) zu 563 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein, um einen Index von .01 – ohne Publikationsbias – zu erreichen.

Aufgrund der Ergebnisse sollte sowohl unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven als auch negativen Wertebereich ein Publikationsbias nicht ausgeschlossen werden. Es sind mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde.

Darüber hinaus ist anhand der Abbildung 20 auffällig, dass im rechten Bereich, außerhalb der simulierten Effektgrößen, vier Effektgrößen publiziert sind. In Tabelle 25 sind zu den vier Effektgrößen studienspezifische Merkmale aufgeführt. Allerdings ließen sich keine eindeutigen Gemeinsamkeiten feststellen, weshalb die publizierten Effektgrößen von den simulierten Effektgrößen im rechten äußeren Bereich abweichen.

**Tabelle 25**

Deskriptive Angaben publizierter Studien, in denen die Effektgröße des Zusammenhangs zwischen  $g$  und  $E$  von den gezogenen Effektgrößen aus der adaptierten Simulation abweicht

| Studie                                   | kogLT | PSK      | Land | Alter                           | Stichprobe                                   | Thema   |
|--|-------|----------|------|---------------------------------|--|---|
| Carson, Stalikas & Bizot (1997)          | BAT   | MBTI     | CAN  | $M = 15.50$<br>( $SD = 0.50$ )  | Schüler<br>( $n = 99$ )                      | Zusammenhang zwischen MBTI und Fähigkeiten                        |
| Parkinson & Taggar (2006)                | GPA   | NEO-PI-R | CAN  | $M = 21.00$<br>( $SD = n.b.$ )  | Studenten<br>( $n = 305$ )                   | Zusammenhang von Intelligenz, Persönlichkeit und Leistung         |
| Batey, Chamorro-Premuzic & Funham (2010) | WPT   | NEO-PI   | GB   | $M = 20.43$<br>( $SD = 2.45$ )  | Studenten<br>( $n = 158$ )                   | Vorhersage von Kreativität durch Persönlichkeit und Intelligenz   |
| Motowidlo, Brownlee & Schmit (2008)      | WPT   | NEO-FFI  | USA  | $M = 26.44$<br>( $SD = 11.63$ ) | Angestellte im Einzelhandel<br>( $n = 140$ ) | Einfluss der Persönlichkeit auf Wissen, Fertigkeiten und Leistung |

*Anmerkungen.* kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; BAT = Ball Aptitude Test; MBTI = Myers Briggs Type Indicator; GPA = Grand Point Average; WPT = Wonderlic Personal Test.

Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durchgeführt wurde:

Nach der „trim and fill“ – Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann aufgrund der Höhe des Schätzers mit  $R_0 = 3$  ein Publikationsbias bei einer mittleren Effektgröße von  $\bar{r} = .01$  nicht ausgeschlossen werden. Nach der anschließenden Exklusion der drei Effektgrößen mit positiven Vorzeichen und den höchsten Beträgen konnte eine symmetrische Verteilung um die neu gebildete mittlere Effektgröße von  $\bar{r} = -.01$  angenommen werden. Der zweite Schätzer  $L_0$  zeigt bei den Effektgrößen mit positiven Vorzeichen kein Publikationsbias an. Allerdings kann bei den Effektgrößen, die kleiner als die mittlere Effektgröße von  $\bar{r} = .01$  sind, ein asymmetrische Verteilung mit  $L_0 = 9.31$  angenommen werden. Durch zwei Iterationen des „trimm“-Schrittes wurden insgesamt 12 publizierte Effektgrößen mit negativen Vorzeichen und den höchsten Beträgen exkludiert. Danach konnte eine Symmetrie mit  $L_0 < 1$  um die neu berechnete mittlere Effektgröße von  $\bar{r} = .03$  angenommen werden. Darüber hinaus wurde auch in dieser Analyse-Variante eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 85 Stichprobengrößen real publizierter Studien und die entsprechende Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 26 aufgeführt.

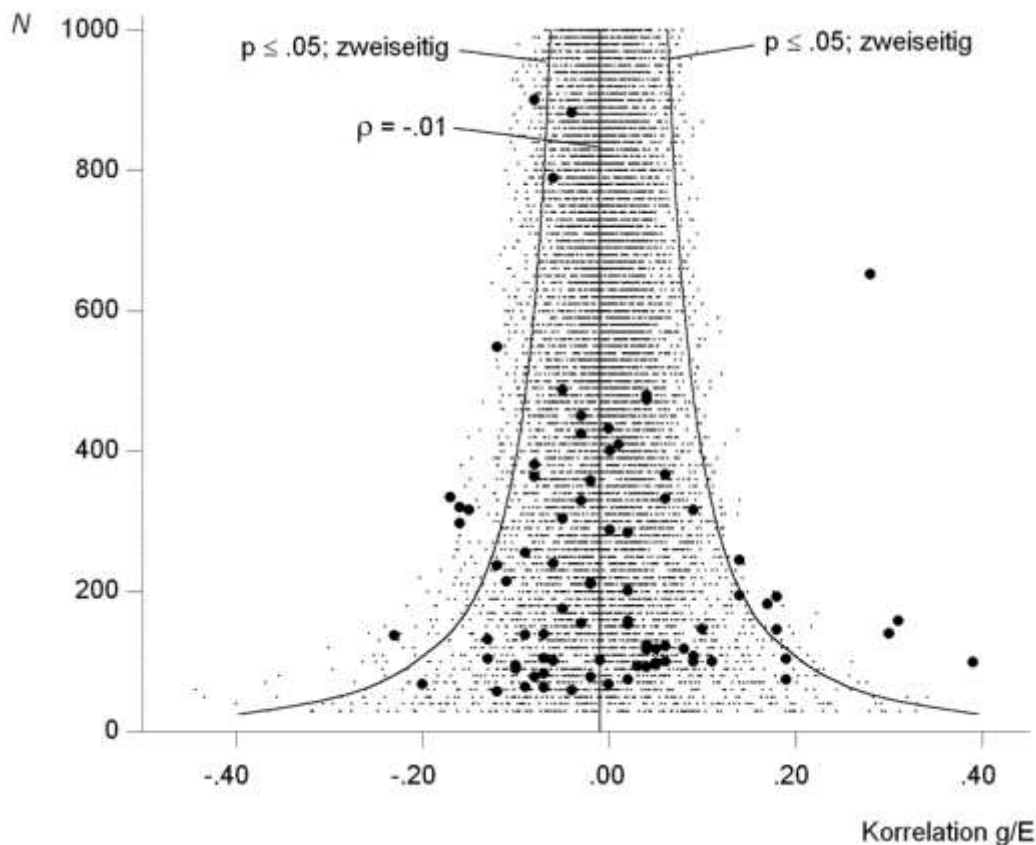
**Tabelle 26**

Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | $N$           | $k$ (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | $k$ (Simulation) <sup>3</sup> |
|-----------|---------------|------------------------------|-----------------------|-------------------------------|
| 1.        | 60 – 100      | 25                           | 29.41                 | 1,471                         |
| 2.        | 110 – 150     | 16                           | 18.82                 | 941                           |
| 3.        | 160 – 200     | 7                            | 8.24                  | 412                           |
| 4.        | 210 – 250     | 6                            | 7.06                  | 353                           |
| 5.        | 260 – 300     | 5                            | 5.88                  | 294                           |
| 6.        | 310 – 350     | 6                            | 7.06                  | 353                           |
| 7.        | 360 – 400     | 5                            | 5.88                  | 294                           |
| 8.        | 410 – 450     | 4                            | 4.71                  | 236                           |
| 9.        | 460 – 500     | 3                            | 3.53                  | 177                           |
| 10.       | 510 – 550     | 1                            | 1.18                  | 59                            |
| 11.       | 610 – 650     | 1                            | 1.18                  | 59                            |
| 12.       | 760 – 800     | 1                            | 1.18                  | 59                            |
| 13.       | 860 – 900     | 2                            | 2.35                  | 118                           |
| 14.       | 1,310 – 1,350 | 1                            | 1.18                  | 59                            |
| 15.       | 1,910 – 1,950 | 1                            | 1.18                  | 59                            |
| 17.       | 4,510 – 4,550 | 1                            | 1.18                  | 59                            |
| $\Sigma$  |               | <b>85</b>                    | <b>100.00</b>         | <b>5,003<sup>4</sup></b>      |

Anmerkungen.  $\rho = -.01$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten  $k$  für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt  $k$  von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt  $k = 5,003$ .

Unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich hat sich aus der adaptierten Simulation (Kapitel 4.6.7) ein Index von .03 berechnet (170 signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 4,833 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig). Bezüglich der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich hat sich aus der Simulation (Kapitel 4.6.7) ein Index von .02 berechnet (78 signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 4,925 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig). Der aus den tatsächlich publizierten Effektgrößen berechnete Index beträgt sowohl unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich als auch unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich .11 (neun signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 76 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 21).



**Abbildung 21.**  $\rho = -.01$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 26) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktionen im positiven und negativen Wertebereich mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie (McGrew, 2009).

Um bei den tatsächlich publizierten Effektgrößen einen Index von .03 – ohne Publikationsbias – zu erreichen, müssten weitere 180 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von neun signifikanten ( $p \leq .05$ , zweiseitig) zu 256 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Um einen Index von .01 zu erreichen, müssten sogar weitere 290 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von neun signifikanten ( $p \leq .05$ , zweiseitig) zu 366 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Dementsprechend kann auch in dieser Analyse-Variante ein Publikationsbias sowohl für den negativen als auch positiven Wertebereich nicht ausgeschlossen werden, da jeweils zu viele signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt sind.

Anhand der Abbildung 21 ist auffällig, dass im rechten Bereich vier Effektgrößen, deutlich außerhalb der simulierten Effektgrößen, publiziert sind. In Tabelle 27 sind zu den vier Effektgrößen studienspezifische Merkmale aufgeführt. Die Studien der Tabelle 27 und 25 sind, bis auf Barrick, Stewart, Neubert und Mount (1998; Tabelle 27) sowie Parkinson und Taggar (2006; Tabelle 25), identisch. Allerdings ließen sich auch hier keine eindeutigen Auffälligkeiten bzw. Gemeinsamkeiten feststellen, weshalb die publizierten Effektgrößen von den simulierten Effektgrößen im rechten äußeren Bereich abweichen.

**Tabelle 27**

Deskriptive Angaben publizierter Studien, in denen die Effektgröße des Zusammenhangs zwischen  $g$  und  $E$  von den gezogenen Effektgrößen aus der adaptierten Simulation abweicht

| Studie                                   | kogLT | PSK     | Land | Alter                           | Stichprobe                                   | Thema   |
|--|-------|---------|------|---------------------------------|--|---|
| Carson, Stalikas & Bizot (1997)          | BAT   | MBTI    | CAN  | $M = 15.50$<br>( $SD = 0.50$ )  | Schüler<br>( $n = 99$ )                      | Zusammenhang zwischen MBTI und Fähigkeiten                        |
| Barrick, Stewart, Neubert & Mount (1998) | WPT   | PCI     | USA  | $M = 41.66$<br>( $SD = 9.59$ )  | Angestellte<br>( $n = 305$ )                 | Teamprozesse und -effektivität                                    |
| Batey, Chamorro-Premuzic & Funham (2010) | WPT   | NEO-PI  | GB   | $M = 20.43$<br>( $SD = 2.45$ )  | Studenten<br>( $n = 158$ )                   | Vorhersage von Kreativität durch Persönlichkeit und Intelligenz   |
| Motowidlo, Brownlee & Schmit (2008)      | WPT   | NEO-FFI | USA  | $M = 26.44$<br>( $SD = 11.63$ ) | Angestellte im Einzelhandel<br>( $n = 140$ ) | Einfluss der Persönlichkeit auf Wissen, Fertigkeiten und Leistung |

*Anmerkungen.* kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; BAT = Ball Aptitude Test; MBTI = Myers Briggs Type Indicator; PCI = Personal Characteristics Inventory; WPT = Wonderlic Personal Test.

## *Kapitel 5 – Ergebnisse*

### *Moderatoranalysen:*

Zum Zusammenhang zwischen g und E beträgt die Varianz des artefaktkorrigierten Effektes ( $\rho$ ) ausnahmslos weniger als 75%. Dementsprechend wurden Moderatoranalysen durchgeführt. Die kognitiven Leistungstests betreffend, wurden die Studien, in denen der WPT zur Erfassung von g eingesetzt wurde mit den Studien verglichen, in denen andere kogLT als der WPT eingesetzt wurden. Des Weiteren wurden Moderatoranalysen hinsichtlich der eingesetzten Persönlichkeitsinventare durchgeführt. Dabei wurden folgende Persönlichkeitsinventare verglichen:

- NEO-Inventare (NEO-PI-(R) oder NEO-FFI) vs.
- NEO-PI-(R) vs.
- NEO-FFI vs.
- FFM-Inventare, allerdings keine NEO-Inventare (bspw. International Personality Item Pool) vs.
- Eysenck-Inventare.

In Tabelle 28 sind die metaanalytischen Kennwerte der entsprechenden Moderatoranalysen zum Zusammenhang zwischen g und E aufgeführt.

**Tabelle 28**  
Moderatoranalysen zum Zusammenhang zwischen g und E

| g/E                     | Moderator   | <i>N</i> | <i>K</i> | $\rho$ | $SD_{\rho}$ | % Var. | 90% CV | $\bar{r}$ | $SD_r$ | % Var  |
|-------------------------|-------------|----------|----------|--------|-------------|--------|--------|-----------|--------|--------|
| <i>kogLT</i>            |             |          |          |        |             |        |        |           |        |        |
| Autor(en) <sup>1</sup>  | WPT         | 4,880    | 28       | .02    | .122        | 36.03  | -.14   | .01       | .102   | 33.76  |
| Autor(en) <sup>1</sup>  | ohne WPT    | 90,093   | 44       | .02    | .057        | 21.02  | -.05   | .02       | .041   | 22.85  |
| Autor(en) <sup>1a</sup> | ohne WPT    | 26,49    | 43       | -.03   | .095        | 26.49  | .09    | -.01      | .083   | 25.66  |
| CHC <sup>2</sup>        | WPT         | 5,532    | 29       | .04    | .166        | 22.91  | -.17   | .04       | .130   | 23.85  |
| CHC <sup>2</sup>        | ohne WPT    | 93,185   | 57       | .02    | .049        | 31.89  | -.04   | .02       | .036   | 32.55  |
| CHC <sup>2a</sup>       | ohne WPT    | 21,298   | 56       | -.03   | .077        | 42.78  | .07    | -.02      | .060   | 42.56  |
| <i>PSK</i>              |             |          |          |        |             |        |        |           |        |        |
| Autor(en) <sup>1</sup>  | FFM, NEO    | 10,159   | 40       | -.04   | .105        | 27.89  | .10    | -.03      | .102   | 27.40  |
| Autor(en) <sup>1</sup>  | NEO-PI-(R)  | 6,892    | 22       | -.05   | .103        | 22.87  | .09    | -.03      | .108   | 21.65  |
| Autor(en) <sup>1</sup>  | NEO-FFI     | 3,267    | 18       | -.01   | .108        | 39.21  | .12    | -.02      | .090   | 40.63  |
| Autor(en) <sup>1</sup>  | FFM, no NEO | 4,224    | 17       | -.03   | .070        | 55.01  | .06    | -.01      | .060   | 53.25  |
| CHC <sup>2</sup>        | FFM, NEO    | 12,105   | 47       | -.02   | .076        | 48.55  | .08    | -.02      | .062   | 50.74  |
| CHC <sup>2</sup>        | NEO-PI-(R)  | 7,212    | 26       | -.02   | .066        | 51.21  | .07    | -.02      | .057   | 52.96  |
| CHC <sup>2</sup>        | NEO-FFI     | 4,891    | 21       | -.03   | .091        | 45.79  | .09    | -.03      | .068   | 48.54  |
| CHC <sup>2</sup>        | FFM, no NEO | 4,441    | 17       | .03    | .175        | 19.02  | -.26   | .05       | .116   | 22.12  |
| CHC <sup>2</sup>        | Eysenck     | 4,205    | 11       | -.13   | .000        | 100.00 | -.13   | -.09      | .000   | 100.00 |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und –überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; % Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein nach der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein nach der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist; kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; WPT = Wonderlic Personal Test (hierbei die Studien inkludiert, die den WPT zur Erfassung der allgemeinen Intelligenzausprägung verwendet haben; ohne WPT = die Studien inkludiert, die zur Erfassung der allgemeinen Intelligenz andere kogLT als den WPT verwendet haben; FFM, NEO = die Studien inkludiert, in denen Costa und McCrae-Inventare (NEO-PI, NEO-PI-R, NEO-FFI) eingesetzt wurden; NEO-PI-(R) = die Studien inkludiert, in denen der NEO-PI-R und der NEO-PI verwendet wurde; NEO-FFI = die Studien berücksichtigt, in denen der NEO-FFI eingesetzt wurde; FFM, no NEO = die Studien berücksichtigt, in denen FFM-Inventare, allerdings keine Costa und McCrae Inventare verwendet wurden; Eysenck = die Studien inkludiert, die E mit Eysenck-Inventaren erfasst haben; grau markiert ist die Zeile, in der mehr als 75% der Varianz erklärt wird.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien erfolgte, konnte durch keine, in der vorliegenden Arbeit durchgeführten Moderatoranalyse mehr als 75% der Varianz zum Zusammenhang zwischen g und E erklärt werden. In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der

CHC-Theorie vorgenommen wurde, wird in den Studien, in denen Eysenck-Inventare eingesetzt wurden, mehr als 75% der Varianz erklärt (Tabelle 28, grau markiert). Die artefaktkorrigierte Effektgröße beträgt in dem Fall  $\rho = -.13$ . Der Unterschied zwischen Eysenck- und Nicht-Eysenck-Inventaren führt zu erheblicher Varianz der Effektstärken. Offenbar sind die konzeptionellen Unterschiede zwischen FFM-Inventaren und Eysenck-Inventaren in Bezug auf Extraversion relevant für die Zusammenhänge mit  $g$ .

*Zusammenfassung zum Zusammenhang zwischen der Allgemeinen Intelligenz und der Extraversion:*

Sowohl in der Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der publizierten Studien berücksichtigt wurde als auch in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) im Nachhinein vorgenommen wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = .02$ . Nach der Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) verringert sich die korrigierte Effektgröße auf  $\rho = -.02$  (Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien) bzw.  $\rho = -.01$  (Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie). Da die erklärte Varianz ausnahmslos kleiner als 75% ist, ist nicht auszuschließen, dass ein mittlerer Effekt heterogener Subpopulationen vorliegt.

Bezüglich eines Publikationsbias kann in der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien vorgenommen wurde, eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .04$  mit  $R_0 < 1$  angenommen werden. Ein Publikationsbias kann demnach ausgeschlossen werden. Der zweite Schätzer  $L_0$  zeigt hingegen eine asymmetrische Verteilung an. Nach zwei Iterationen, in denen insgesamt 18 „trim“-Schritte im negativen Wertebereich der publizierten Effektgrößen durchgeführt wurden, kann eine symmetrische Verteilung der Effektgrößen hinsichtlich der Summe aller Ränge um die mittlere Effektgröße von  $\bar{r} = .05$  mit  $L_0 < 1$  angenommen und ein Publikationsbias ausgeschlossen werden.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, zeigten beide Schätzer  $R_0$  und  $L_0$  eine asymmetrische Verteilung der publizierten Effektgrößen an. Ein Publikationsbias ist demnach nicht auszuschließen. Durch die Exklusion von drei Effektgrößen mit den höchsten Beträgen im positiven Wertebereich kann eine symmetrische Verteilung um die neu gebildete mittlere Effektgröße von  $\bar{r} = -.01$  mit  $R_0 < 1$  angenommen werden. Der zweite Schätzer  $L_0$  zeigt bei den



publizierten Effektgrößen im negativen Wertebereich eine asymmetrische Verteilung bezüglich der Summe aller Ränge an. Durch zwei Iterationen, in denen insgesamt 12 Effektgrößen mit den höchsten Beträgen im negativen Wertebereich exkludiert wurden, kann eine Symmetrie hinsichtlich der Summe aller Ränge um die die mittlere neu berechnete Effektgröße von  $\bar{r} = .03$  mit  $L_0 < 1$  angenommen und ein Publikation ausgeschlossen werden.

Nach den durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens, kann für beide Analyse-Varianten (Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien vs. CHC-Theorie) ein Publikationsbias nicht ausgeschlossen werden, da zu viele signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt sind. Dabei kann von einem Publikationsbias im positiven und negativen Wertebereich ausgegangen werden.

Hinsichtlich der Moderatoranalysen konnte bei den Studien, die zur Erfassung von E Eysenck-Inventar verwendeten, mehr als 75% der Varianz erklärt werden (Tabelle 28). Offenbar sind die konzeptionellen Unterschiede zwischen dem FFM und Eysenck in Bezug auf Extraversion relevant für die Zusammenhänge mit  $g$ .

### 5.5 Ergebnisse zur Fluiden Intelligenz und Extraversion

#### *Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien:*

In dieser Analyse-Variante beträgt die inkludierte Studienanzahl  $k = 26$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .01$  ( $SD = .03$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 1,052$  (Range:  $n = 57$  bis  $20,175$ ;  $SD = 3,934.92$ ;  $N = 27,355$ :). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. -überschätzung korrigierte Effektgröße beträgt  $\rho = .00$  ( $SD_\rho = .04$ ). Die erklärte Varianz beträgt 47.90% (Tabelle 29).

Durch die anschließende Exklusion der Studie von De Fruyt et al. (2006,  $n = 20,175$ ) verringerte sich die unkorrigierte Effektgröße von  $\bar{r}_{bare-bones} = .01$  ( $SD = .03$ ) auf  $\bar{r}_{bare-bones} = -.03$  ( $SD = .03$ ). Die mittlere Stichprobengröße beträgt hierbei  $\bar{n} = 287$  (Range:  $n = 57$  bis  $2,658$ ;  $SD = 518.82$ ;  $N = 7,172$ ). Der artefaktkorrigierte Effektgröße veränderte sich von  $\rho = .00$  ( $SD_\rho = .04$ ) auf  $\rho = -.02$  ( $SD_\rho = .03$ ). Die Null liegt hierbei im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 76.01% (Tabelle 29).

Offenbar wird der Zusammenhang zwischen Gf und E in dieser Analyse-Variante von einer einzigen Studie (De Fruyt et al., 2006) mit einer Stichprobengröße von  $n = 20,175$  moderiert. Nach der Exklusion der Studie von De Fruyt et al. (2006) können mehr als 75% der Varianz erklärt werden. Dementsprechend wurden in dieser Analyse-Variante keine Moderatoranalysen durchgeführt. Aufgrund der Ergebnisse kann unabhängig von der Exklusion bzw. Inklusion der Studie von De Fruyt et al. (2006) ein Populationszusammenhang von Null nicht ausgeschlossen werden.

**Tabelle 29**

Metaanalytische Ergebnisse zum Zusammenhang zwischen Gf und E

| Gf/Extraversion                      | $N$    | $k$ | $\rho$ | $SD_{\rho}$ | %Var. | 90% CV | $\bar{r}$ | $SD_r$ | %Var  |
|--------------------------------------|--------|-----|--------|-------------|-------|--------|-----------|--------|-------|
| kodiert nach Autor(en) <sup>1</sup>  | 27,355 | 26  | .00    | .035        | 57.72 | -.04   | .01       | .025   | 59.58 |
| kodiert nach Autor(en) <sup>1a</sup> | 7,172  | 25  | -.02   | .032        | 76.01 | .02    | -.03      | .031   | 78.47 |
| kodiert nach CHC <sup>2</sup>        | 40,298 | 60  | .00    | .049        | 50.10 | -.07   | .00       | .039   | 50.06 |
| kodiert nach CHC <sup>2a</sup>       | 20,115 | 59  | -.02   | .052        | 55.95 | .05    | -.02      | .047   | 57.18 |

*Anmerkungen.*  $N$  = Stichprobengröße;  $k$  = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und –überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; %Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße; <sup>1</sup> = Datensätze, in denen eine Zuordnung von Gf explizit von den Autoren der publizierten Studien vorgenommen wurde („Bare-Bones“-Analyse); <sup>1a</sup> = Datensätze, in denen eine Zuordnung von Gf explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von De Fruyt et al. (2006,  $n = 20,175$ ) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein nach der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein nach der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde und die Studie von De Fruyt et al. (2006,  $n = 20,175$ ) exkludiert ist.

*Zuordnung der Fähigkeitskonstrukte gemäß CHC-Theorie:*

In dieser Analyse-Variante beträgt die inkludierte Studienzahl  $k = 60$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .00$  ( $SD_{bare-bones} = .04$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 672$  (Range:  $n = 57$  bis  $20,175$ ;  $SD = 2,619.40$ ;  $N = 40,298$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. –überschätzung korrigierte Effektgröße beträgt  $\rho = .00$  ( $SD_{\rho} = .05$ ). Die erklärte Varianz beträgt 50.10% (Tabelle 29).

Durch die anschließende Exklusion der Studie von De Fruyt et al. (2006,  $n = 20,175$ ), verringert sich die unkorrigierte Effektgröße von  $\bar{r}_{bare-bones} = .00$  ( $SD = .04$ ) auf  $\bar{r}_{bare-bones} = -.02$  ( $SD_{\rho} = .05$ ). Die mittlere Stichprobengröße beträgt hierbei  $\bar{n} = 341$  (Range:  $n = 57$  bis  $2,982$ ;  $SD = 551.92$ ;  $N = 20,115$ ). Der Betrag der artefaktkorrigierten Effektgröße verringerte sich von  $\rho = .00$  ( $SD_{\rho} = .05$ ) auf  $\rho = -.02$  ( $SD_{\rho} = .05$ ). Die Null liegt im Glaubwürdigkeitsintervall und die erklärte Varianz

beträgt 55.95%. Unabhängig von der Exklusion bzw. Inklusion der Studie von De Fruyt et al. (2006) kann auch in dieser Analyse-Variante ein Populationszusammenhang von Null nicht ausgeschlossen werden. Aufgrund einer erklärten Varianz, die deutlich kleiner als 75% ist, kann davon ausgegangen werden, dass die artefaktkorrigierte Effektgröße ein mittlerer Effekt heterogener Subpopulationen ist. Dementsprechend wurden in dieser Analyse-Variante Moderatoranalysen durchgeführt (siehe unten).

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe der Schätzer mit  $R_0 < 1$  und  $L_0 < 1$  eine weitestgehend symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.02$  angenommen werden. Demnach ist ein Publikationsbias auszuschließen. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 25 Stichprobengrößen real publizierter Studien und die entsprechende Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 30 aufgeführt.

**Tabelle 30**

Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

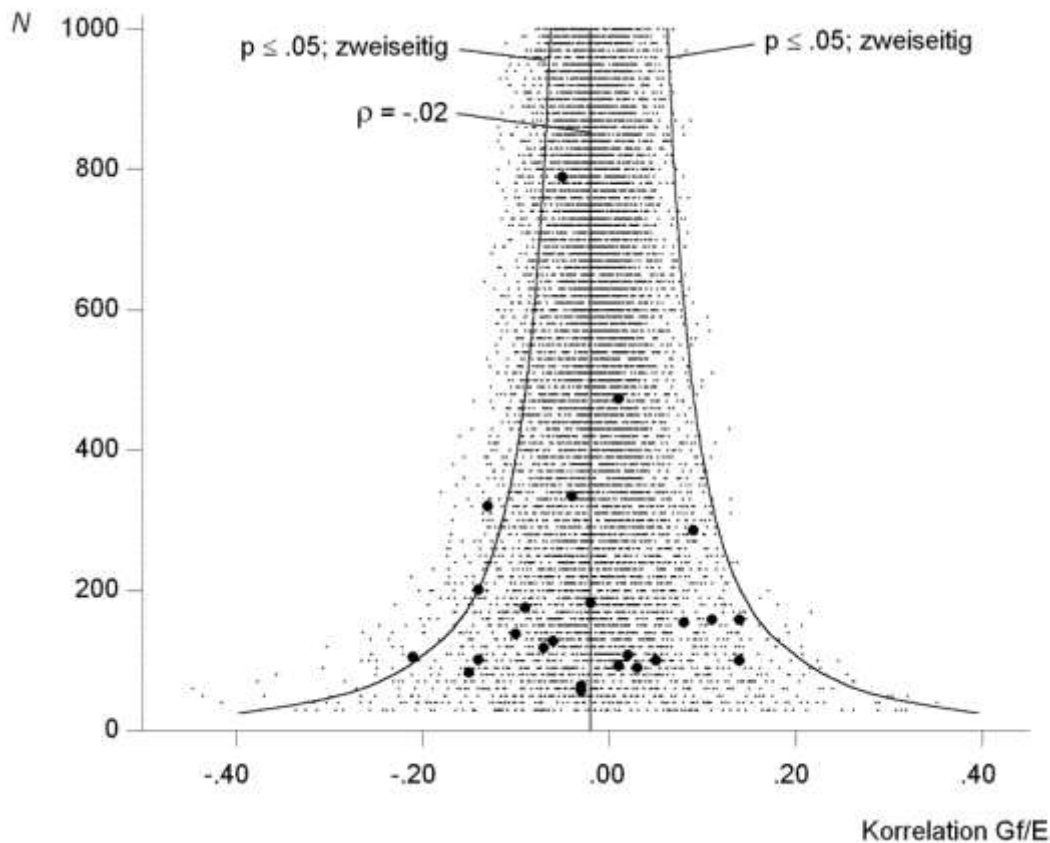
| Intervall | <i>N</i>      | <i>k</i> (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | <i>k</i> (Simulation) <sup>3</sup> |
|-----------|---------------|-----------------------------------|-----------------------|------------------------------------|
| 1.        | 60 – 100      | 8                                 | 32.00                 | 1,600                              |
| 2.        | 110 – 150     | 6                                 | 24.00                 | 1,200                              |
| 3.        | 160 – 200     | 5                                 | 20.00                 | 1,000                              |
| 4.        | 260 – 300     | 1                                 | 4.00                  | 200                                |
| 5.        | 310 – 350     | 2                                 | 8.00                  | 400                                |
| 6.        | 460 – 500     | 1                                 | 4.00                  | 200                                |
| 7.        | 760 – 800     | 1                                 | 4.00                  | 200                                |
| 8.        | 2,660 – 2,700 | 1                                 | 4.00                  | 200                                |
| <b>Σ</b>  |               | <b>25</b>                         | <b>100.00</b>         | <b>5,000</b>                       |

Anmerkungen.  $\rho = -.02$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten *k* für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt *k* von 5,000.

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .04 unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich hat sich aus einem Verhältnis von 222 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,778 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Da die simulierten Effektgrößen keinerlei Publikationsbias aufweisen, kann hierbei ein Verhältnis von 4% signifikanten ( $p \leq .05$ , zweiseitig)

zu 96% nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) als Referenz zur Beurteilung eines Publikationsbias herangezogen werden.

Da keine der tatsächlich publizierten Effektgrößen mit positiven Vorzeichen signifikant ist, kann in dem Fall ein Publikationsbias ausgeschlossen werden. Deshalb wurde keine Simulation unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich durchgeführt (Abbildung 22). Die Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich berücksichtigend, beträgt der Index aus den tatsächlich publizierten Effektgrößen .12 (drei signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 22 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 22).



**Abbildung 22.**  $\rho = -.02$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 30) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktionen im positiven und negativen Wertebereich mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien.

Um bei den tatsächlich publizierten Effektgrößen einen Index von .04 – ohne Publikationsbias – zu erreichen, müssten unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im

negativen Wertebereich weitere 43 nichtsignifikante Effektgrößen ( $p \leq .05$ , zweiseitig) in einem Verhältnis von drei signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 65 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Dementsprechend kann ein Publikationsbias, bei dem zu viele signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt sind, nicht ausgeschlossen werden.

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durchgeführt wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe der Schätzer mit  $R_0 < 1$  und  $L_0 < 1$  eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.02$  angenommen und dementsprechend ein Publikationsbias ausgeschlossen werden. Auch in dieser Analyse-Variante wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 59 Stichprobengrößen real publizierter Studien und die entsprechende Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 31 aufgeführt.

**Tabelle 31**

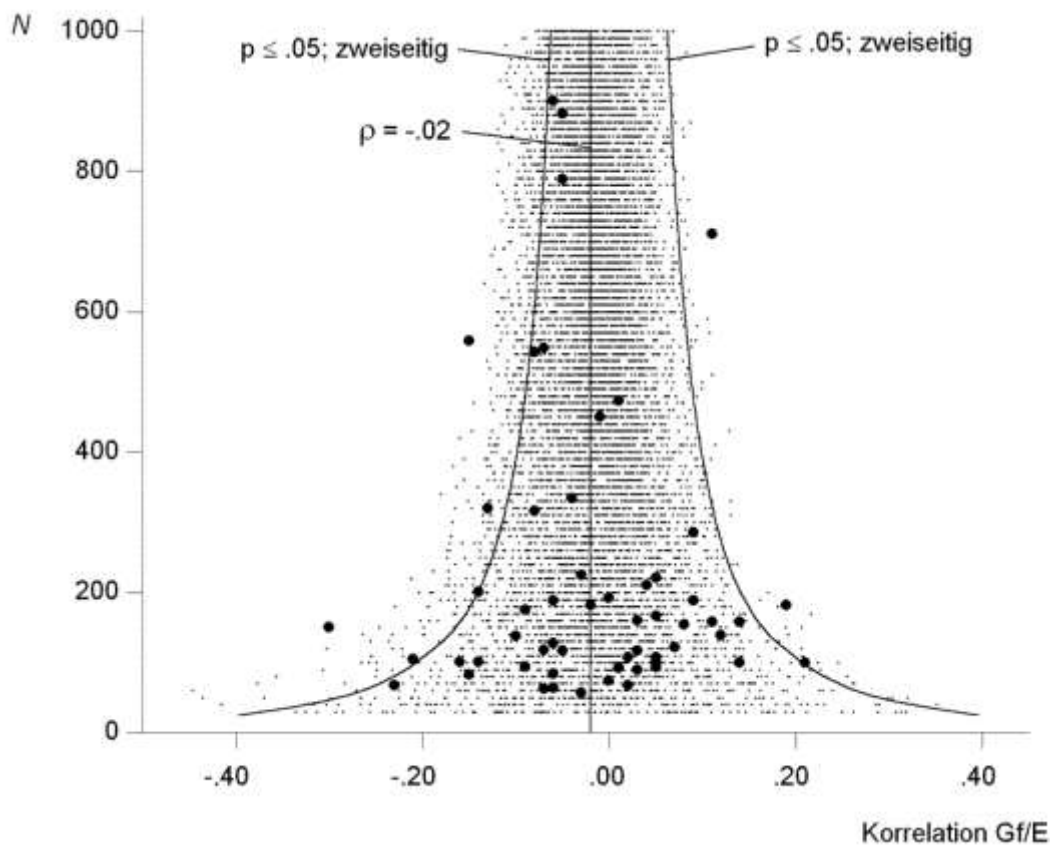
Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | <i>N</i>      | <i>k</i> (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | <i>k</i> (Simulation) <sup>3</sup> |
|-----------|---------------|-----------------------------------|-----------------------|------------------------------------|
| 1.        | 60 – 100      | 17                                | 28.81                 | 1,441                              |
| 2.        | 110 – 150     | 12                                | 20.34                 | 1,017                              |
| 3.        | 160 – 200     | 11                                | 18.64                 | 932                                |
| 4.        | 210 – 250     | 3                                 | 5.08                  | 254                                |
| 5.        | 260 – 300     | 1                                 | 1.70                  | 85                                 |
| 6.        | 310 – 350     | 3                                 | 5.08                  | 254                                |
| 7.        | 410 – 450     | 1                                 | 1.70                  | 85                                 |
| 8.        | 460 – 500     | 1                                 | 1.70                  | 85                                 |
| 9.        | 510 – 550     | 2                                 | 3.39                  | 170                                |
| 10.       | 560 – 600     | 1                                 | 1.70                  | 85                                 |
| 11.       | 710 – 750     | 1                                 | 1.70                  | 85                                 |
| 12.       | 760 – 800     | 1                                 | 1.70                  | 85                                 |
| 13.       | 860 – 900     | 2                                 | 3.39                  | 170                                |
| 14.       | 1,810 – 1,850 | 1                                 | 1.70                  | 85                                 |
| 15.       | 2,660 – 2,700 | 1                                 | 1.70                  | 85                                 |
| 16.       | 2,960 – 3,000 | 1                                 | 1.70                  | 85                                 |
| <b>Σ</b>  |               | <b>59</b>                         | <b>100.00</b>         | <b>5,003<sup>4</sup></b>           |

*Anmerkungen.*  $\rho = -.02$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten *k* für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt *k* von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt *k* = 5,003.

Der aus der adaptierten Simulation gebildete Index von .05 unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich hat sich aus einem Verhältnis von 231 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,772 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Hinsichtlich der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich hat sich ein Index aus der Simulation von .01 aus einem Verhältnis von 52 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,951 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet.

Der aus den tatsächlich publizierten Effektgrößen berechnete Index unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich beträgt .08 (fünf signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 54 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 23). Hinsichtlich der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich beträgt der Index aus den tatsächlich publizierten Effektgrößen .05 (drei signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 56 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 23).



**Abbildung 23.**  $\rho = -.02$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 31) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktionen im positiven und negativen Wertebereich mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie (McGrew, 2009).

Um einen Index von .05 zu erreichen, müssten unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich 34 weitere nichtsignifikante Effektgrößen in einem Verhältnis von fünf signifikanten ( $p \leq .05$ , zweiseitig) zu 88 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Hinsichtlich der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich müssten sogar weitere 155 nichtsignifikante Effektgrößen in einem Verhältnis von drei signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 211 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein, um einen Index von .01 zu erreichen. Dementsprechend kann in dieser Analyse-Variante ein Publikationsbias für den negativen und positiven Wertebereich nicht ausgeschlossen werden. Offenbar sind mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Vergleich zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde.

*Moderatoranalysen:*

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde, konnten, nachdem die Studie von De Fruyt et al. (2006;  $n = 20,175$ ) exkludiert wurde, mehr als 75% der Varianz erklärt werden. Dementsprechend wurden in dieser Analyse-Variante keine Moderatoranalysen durchgeführt.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, beträgt die erklärte Varianz – unabhängig von der Exklusion bzw. Inklusion von De Fruyt et al. (2006) – deutlich weniger als 75%. Dementsprechend wurden in dieser Analyse-Variante Moderatoranalysen durchgeführt. Die kognitiven Leistungstests berücksichtigend, wurden die Studien, in denen der BRT zur Erfassung der fluiden Intelligenzausprägung eingesetzt wurde, mit den Studien verglichen, in denen andere kogLT als der BRT eingesetzt wurde. Weiterhin wurden die Studien, in denen Raven Tests eingesetzt wurden, mit Studien verglichen, in denen andere kogLT als die Raven Tests verwendet wurden. Darüber hinaus wurden Moderatoranalysen hinsichtlich der eingesetzten Persönlichkeitsinventare durchgeführt. Dabei wurden folgende Persönlichkeitsinventare verglichen:

- FFM-Inventare (bspw. NEO-PI-(R), NEO-FFI, International Personality Item Pool)
- NEO-Inventare (NEO-PI-(R) oder NEO-FFI) vs.
- NEO-PI-(R) vs.
- NEO-FFI vs.

## Kapitel 5 – Ergebnisse

- FFM-Inventare, allerdings keine NEO-Inventare vs.
- Eysenck-Inventare.

In Tabelle 32 sind die metaanalytischen Kennwerte der entsprechenden Moderatoranalysen zum Zusammenhang zwischen Gf und E aufgeführt.

**Tabelle 32**  
Moderatoranalysen zum Zusammenhang zwischen Gf und E

| Gf/E              | Moderator  | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | %Var.  | 90% CV | $\bar{r}$ | $SD_r$ | %Var   |
|-------------------|------------|----------|----------|--------|-------------|--------|--------|-----------|--------|--------|
| <i>kogLT</i>      |            |          |          |        |             |        |        |           |        |        |
| CHC <sup>1</sup>  | BRT        | 1,745    | 13       | -.01   | .048        | 80.68  | .04    | -.02      | .044   | 79.62  |
| CHC <sup>1</sup>  | ohne BRT   | 38,553   | 47       | .00    | .049        | 45.59  | -.06   | .00       | .038   | 45.56  |
| CHC <sup>1a</sup> | ohne BRT   | 18,378   | 46       | -.02   | .057        | 47.07  | .06    | -.02      | .052   | 47.70  |
| CHC <sup>1</sup>  | Raven      | 23,881   | 15       | .03    | .032        | 59.17  | -.02   | .02       | .020   | 60.15  |
| CHC <sup>1a</sup> | Raven      | 3,706    | 14       | .00    | .067        | 59.03  | -.09   | .00       | .051   | 59.22  |
| CHC <sup>1</sup>  | ohne Raven | 16,417   | 45       | -.02   | .049        | 55.73  | .04    | -.03      | .045   | 57.97  |
| <i>PSK</i>        |            |          |          |        |             |        |        |           |        |        |
| CHC <sup>1</sup>  | FFM        | 35,198   | 43       | .00    | .048        | 45.55  | -.06   | .00       | .038   | 45.56  |
| CHC <sup>1a</sup> | FFM        | 15,023   | 42       | -.02   | .053        | 51.90  | .048   | -.02      | .049   | 53.63  |
| CHC <sup>1</sup>  | NEO        | 33,517   | 38       | .00    | .046        | 46.70  | -.06   | .00       | .036   | 46.61  |
| CHC <sup>1a</sup> | NEO        | 13,342   | 37       | -.02   | .048        | 56.22  | .04    | -.03      | .044   | 58.62  |
| CHC <sup>1</sup>  | NEO-PI-(R) | 30,288   | 24       | .00    | .045        | 38.54  | -.06   | .01       | .035   | 38.90  |
| CHC <sup>1a</sup> | NEO-PI-(R) | 10,113   | 23       | -.02   | .051        | 45.65  | .05    | -.02      | .051   | 46.43  |
| CHC <sup>1</sup>  | NEO-FFI    | 3,229    | 14       | -.06   | .000        | 100.00 | -.06   | -.05      | .000   | 100.00 |
| CHC <sup>1</sup>  | Eysenck    | 1,550    | 8        | .04    | .055        | 73.66  | -.03   | .03       | .041   | 75.36  |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und -überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; %Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde; <sup>1a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde und die Studie von De Fruyt et al. (2006, *n* = 20,175) exkludiert ist; kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; BRT = Baddeley Reasoning Test (hierbei die Studien inkludiert, die den BRT zur Erfassung der fluiden Intelligenzausprägung verwendet haben; ohne BRT = die Studien inkludiert, die zur Erfassung der fluiden Intelligenz andere kogLT als den BRT verwendet haben; FFM = die Studien inkludiert, in denen „Big-Five“-Inventare eingesetzt wurden; NEO = die Studien inkludiert, in denen ausschließlich Costa und McCrae-Inventare (NEO-PI, NEO-PI-R, NEO-FFI) eingesetzt wurden; NEO-PI-(R) = die Studien inkludiert, in denen der NEO-PI-R und der NEO-PI verwendet wurde; NEO-FFI = die Studien berücksichtigt, in denen der NEO-FFI eingesetzt wurde; Eysenck = die Studien inkludiert, die E mit Eysenck-Inventaren erfasst haben; grau markiert sind die Zeilen, in denen mehr als 75% der Varianz erklärt wird.



In den Studien, in denen der BRT zur Messung von Gf eingesetzt wurde, werden mehr als 75% der Varianz erklärt (Tabelle 32, grau markiert). Offenbar sind die konzeptionellen Unterschiede zwischen dem BRT-Tests und Nicht-BRT-Tests in Bezug auf Gf relevant für die Zusammenhänge mit E. Dabei kann von einem Nulleffekt in dieser Population ausgegangen werden. Weiterhin wird in den Studien, in denen der NEO-FFI zur Erfassung von E verwendet wurde, mehr als 75% der Varianz erklärt. Dabei kann von einem Zusammenhang von  $\rho = -.06$  für diese Subpopulation ausgegangen werden (Tabelle 32, grau markiert). Die konzeptionellen Unterschiede zwischen dem NEO-FFI und den Nicht-NEO-FFI-Inventaren scheinen in Bezug auf E relevant für die Zusammenhänge mit Gf zu sein.

*Zusammenfassung zum Zusammenhang zwischen der Fluiden Intelligenz und der Extraversion:*

Sowohl on der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der publizierten Studien beibehalten wurde als auch in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = .00$ . Nach der Exklusion der Studie von De Fruyt et al. (2006) mit einer Stichprobengröße von  $n = 71,887$  verringert sich die korrigierte Effektgröße in beiden Analyse-Varianten auf  $\rho = -.02$ . Ein Populationseffekt von Null ist nicht auszuschließen. I

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien unverändert blieb, konnten nach der Exklusion der Studie von De Fruyt et al. (2006,  $n = 71,887$ ) mehr als 75% der Varianz erklärt werden. Dementsprechend ist davon auszugehen, dass der Zusammenhang zwischen Gf und E in dieser Analyse-Variante durch eine einzelne Studie mit einer sehr großen Stichprobengröße moderiert wird.

Hinsichtlich eines Publikationsbias kann in beiden Analyse-Varianten eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße angenommen und dementsprechend ein Publikationsbias ausgeschlossen werden. Nach den durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens, kann in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien unverändert blieb, ein Publikationsbias unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich nicht ausgeschlossen werden. Dabei sind zu viele signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert als man bei einer zufälligen Verteilung signifikanter ( $p \leq .05$ , zweiseitig) und nichtsignifikanter Effektgrößen ( $p > .05$ , zweiseitig) erwarten würde. In der

Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, kann sowohl im positiven als auch negativen Wertebereich ein Publikationsbias nicht ausgeschlossen werden, da jeweils zu viele signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt sind.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, konnten in den Studien, in denen der BRT oder NEO-FFI verwendet wurde, jeweils mehr als 75% der Varianz erklärt werden. Dementsprechend ist davon auszugehen, dass die konzeptionellen Unterschiede zwischen dem BRT-Tests und Nicht-BRT-Tests bzw. NEO-FFI und Nicht-NEO-FFI in Bezug auf den Zusammenhang zwischen Gf und E relevant sind.

## 5.6 Ergebnisse zur Kristallisierten Intelligenz und Extraversion

*Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien:*

In der Variante beträgt die inkludierte Studienanzahl  $k = 12$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{\text{bare-bones}} = .06$  ( $SD = .03$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 6,360$  (Range:  $n = 57$  bis  $71,887$ ;  $SD = 20,641.64$ ;  $N = 76,316$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. -überschätzung korrigierte Effektgröße beträgt  $\rho = .09$  ( $SD_{\rho} = .04$ ). Das Glaubwürdigkeitsintervall („credibility interval“) schließt die Null nicht mit ein und die erklärte Varianz beträgt 15.17%. (Tabelle 33).

Nach der anschließenden Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) verringert sich die unkorrigierte Effektgröße von  $\bar{r}_{\text{bare-bones}} = .06$  ( $SD = .03$ ) auf  $\bar{r}_{\text{bare-bones}} = -.05$  ( $SD = .04$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 403$  (Range:  $n = 57$  bis  $1,821$ ;  $SD = 516.86$ ;  $N = 4,429$ ). Die artefaktkorrigierte Effektgröße verringert sich hierbei von  $\rho = .09$  ( $SD_{\rho} = .04$ ) auf  $\rho = -.07$  ( $SD_{\rho} = .06$ ). Das Glaubwürdigkeitsintervall schließt die Null nach der Exklusion von Reeve et al. (2006) mit ein und die erklärte Varianz beträgt 57.31% (Tabelle 33).

Es kann in dieser Analyse-Variante von einem positiven Populationszusammenhang ausgegangen werden, solange die Studie von Reeve et al. (2006,  $n = 71,887$ ) nicht exkludiert wird. Die Höhe der artefaktkorrigierten positiven Effektgröße sollte nicht generalisiert werden, da aufgrund der erklärten Varianz  $< 75\%$  die Höhe der artefaktkorrigierten Effektgröße von Subpopulationen moderiert wird. Dementsprechend wurden Moderatoranalysen durchgeführt (siehe unten). Durch die anschließende Exklusion von Reeve et al. (2006,  $n = 71,887$ ) kann in der Population eine

Effektgröße von Null nicht mehr ausgeschlossen werden, da das Glaubwürdigkeitsintervall die Null mit einschließt. Die erklärte Varianz nimmt um 42.14% auf 57.31% deutlich zu, was als Hinweis gelten kann, dass die Studie von Reeve et al. (2006) einen großen Einfluss auf die metaanalytischen Kennwerte hat.

**Tabelle 33**

Metaanalytische Ergebnisse zum Zusammenhang zwischen Gc und E

| Gc/Extraversion                      | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | %Var. | 90% CV | $\bar{r}$ | $SD_r$ | %Var  |
|--------------------------------------|----------|----------|--------|-------------|-------|--------|-----------|--------|-------|
| kodiert nach Autor(en) <sup>1</sup>  | 76,316   | 12       | .09    | .041        | 15.17 | .04    | .06       | .030   | 14.88 |
| kodiert nach Autor(en) <sup>1a</sup> | 4,429    | 11       | -.07   | .061        | 57.31 | .01    | -.05      | .041   | 59.17 |
| kodiert nach CHC <sup>2</sup>        | 79,948   | 33       | .07    | .061        | 16.77 | .00    | .06       | .043   | 18.25 |
| kodiert nach CHC <sup>2a</sup>       | 8,061    | 32       | -.07   | .070        | 52.94 | .02    | -.06      | .057   | 55.37 |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und –überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; %Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analyse); <sup>1</sup> = Datensätze, in denen eine Zuordnung von Gc explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von Gc explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist.

*Zuordnung der Fähigkeitskonstrukte gemäß CHC-Theorie:*

In dieser Analyse-Variante beträgt die inkludierte Studienzahl *k* = 33 bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .06$  ( $SD = .04$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 2,423$  (Range: *n* = 57 bis 71,887;  $SD = 12,471.86$ ; *N* = 79,948). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. –überschätzung korrigierte Effektgröße beträgt  $\rho = .07$  ( $SD_{\rho} = .06$ ). Die erklärte Varianz beträgt 16.77% und das Glaubwürdigkeitsintervall schließt die Null mit ein.

Nach der anschließenden Exklusion der Studie von Reeve et al. (2006, *n* = 71,887) verringert sich die unkorrigierte Effektgröße von  $\bar{r}_{bare-bones} = .06$  ( $SD = .04$ ) auf  $\bar{r}_{bare-bones} = -.06$  ( $SD = .06$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 252$  (*n* = Range: 57 bis 882;  $SD = 214.44$ ; *N* = 8,061). Der Betrag der artefaktkorrigierten Effektgröße verringert sich hierbei von  $\rho = .05$  ( $SD_{\rho} = .06$ ) auf  $\rho = -.07$  ( $SD_{\rho} = .07$ ). Die Null liegt hierbei ebenfalls nicht im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 52.94% (Tabelle 33).

## Kapitel 5 – Ergebnisse

In dieser Analyse-Variante kann unabhängig von der Exklusion bzw. Inklusion der Studie von Reeve et al. (2006) ein Nulleffekt in der Population nicht ausgeschlossen werden. Aufgrund der erklärten Varianz, die in dieser Analyse-Variante ebenfalls deutlich kleiner als 75% ist, wurde Moderatoranalysen durchgeführt (siehe unten).

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien vorgenommen wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe der Schätzer mit  $R_0 < 1$  und  $L_0 < 1$  eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.07$  angenommen und demnach ein Publikationsbias ausgeschlossen werden. Weiterhin wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 11 Stichprobengrößen real publizierter Studien und die entsprechende Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 34 aufgeführt.

**Tabelle 34**

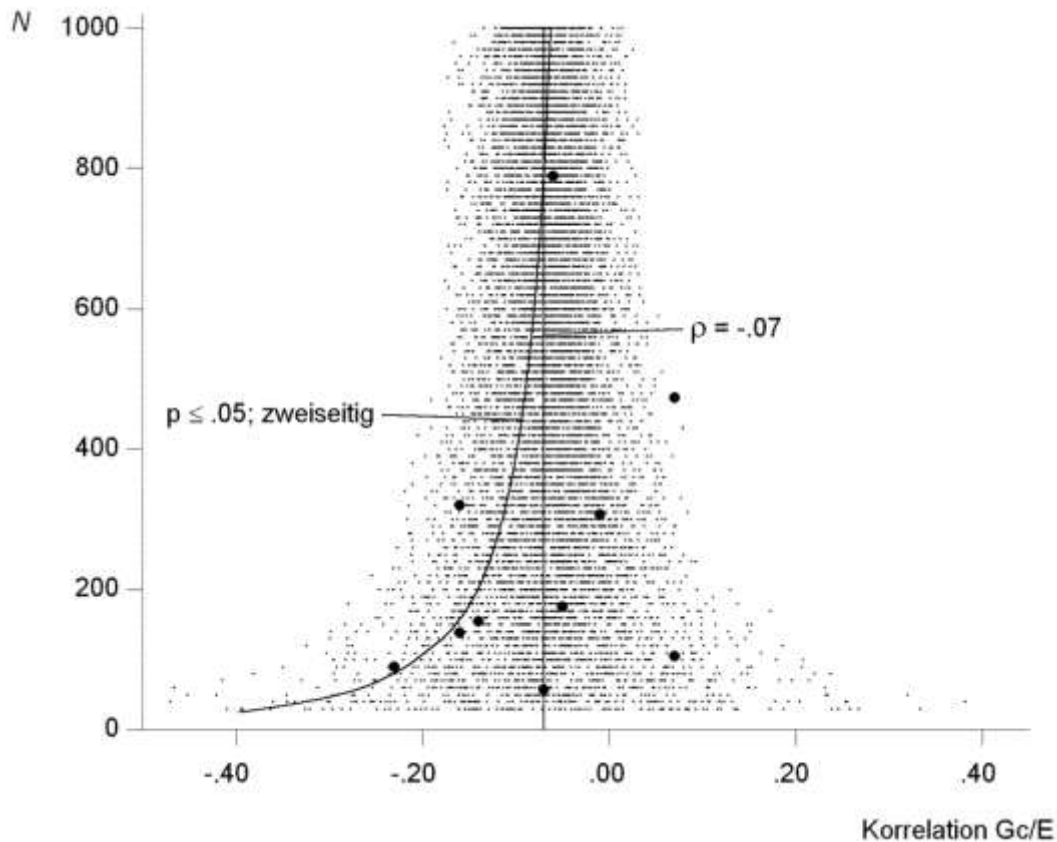
Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | $N$           | $k$ (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | $k$ (Simulation) <sup>3</sup> |
|-----------|---------------|------------------------------|-----------------------|-------------------------------|
| 1.        | 60 – 100      | 2                            | 18.18                 | 909                           |
| 2.        | 110 – 150     | 3                            | 27.27                 | 1,364                         |
| 3.        | 160 – 200     | 1                            | 9.09                  | 455                           |
| 4.        | 310 – 350     | 2                            | 18.18                 | 909                           |
| 5.        | 460 – 500     | 1                            | 9.09                  | 455                           |
| 6.        | 760 – 800     | 1                            | 9.09                  | 455                           |
| 7.        | 1,810 – 1,850 | 1                            | 9.09                  | 455                           |
| $\Sigma$  |               | <b>11</b>                    | <b>100.00</b>         | <b>5,002<sup>4</sup></b>      |

Anmerkungen.  $\rho = -.07$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten  $k$  für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt  $k$  von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt  $k = 5,002$ .

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .23 hat sich aus einem Verhältnis von 1,152 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 3,850 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Da die simulierten Effektgrößen keinerlei Publikationsbias aufweisen, kann hierbei ein Verhältnis von 23% signifikanten ( $p \leq .05$ , zweiseitig) zu 77% nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) als Referenz zur Beurteilung eines Publikationsbias herangezogen werden. Der aus den tatsächlich publizierten

Effektgrößen berechnete Index beträgt .27 (drei signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; acht nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 24).



**Abbildung 24.**  $\rho = -.07$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 34) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien.

Um bei den tatsächlich publizierten Effektgrößen einen Index von .23 – ohne Publikationsbias – zu erreichen, müssten zwei weitere nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von drei signifikanten ( $p \leq .05$ , zweiseitig) zu 10 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) unter den gegebenen Parametern publiziert sein. Dementsprechend kann tendenziell ein Publikationsbias zum gegenwärtigen Zeitpunkt nicht ausgeschlossen werden.

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durchgeführt wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) kann auf Grund der Höhe der Schätzer mit  $R_0 < 1$  und  $L_0 < 1$  eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.05$  angenommen und ein Publikationsbias

demnach ausgeschlossen werden. Darüber hinaus wurde auch hier eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 32 Stichprobengrößen real publizierter Studien und die entsprechende Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 35 aufgeführt.

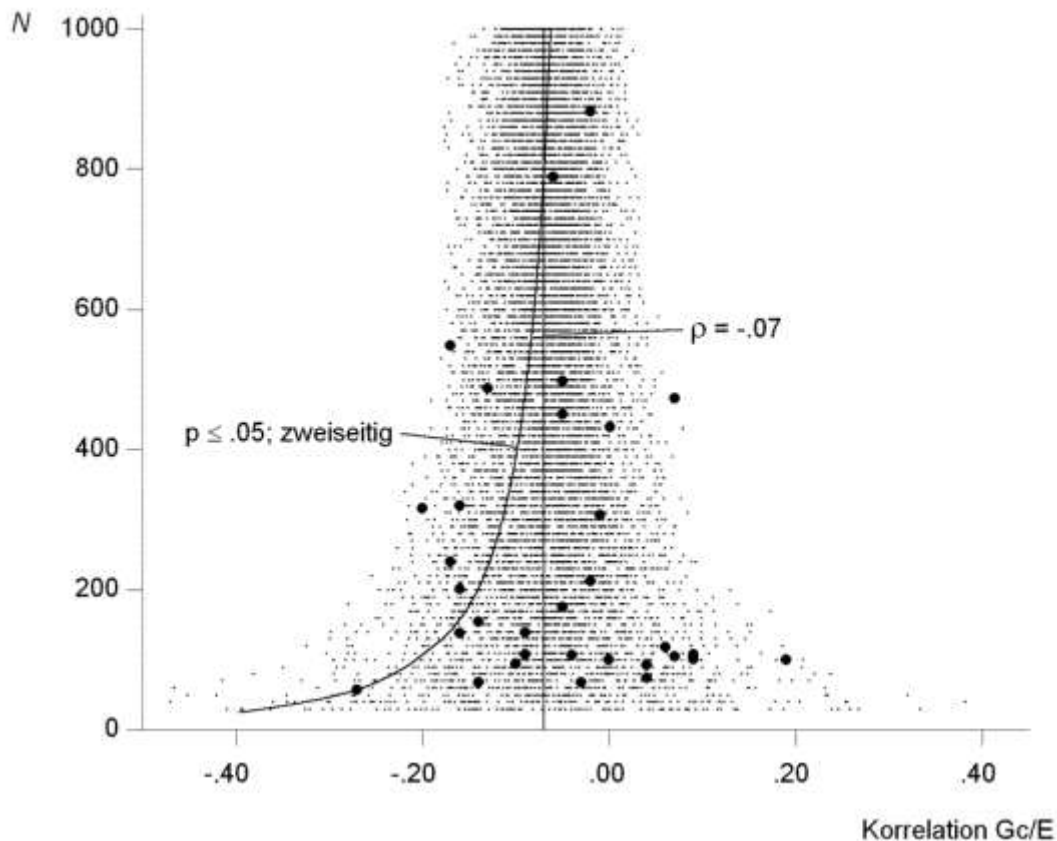
**Tabelle 35**

Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | <i>N</i>  | <i>k</i> (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | <i>k</i> (Simulation) <sup>3</sup> |
|-----------|-----------|-----------------------------------|-----------------------|------------------------------------|
| 1.        | 60 – 100  | 9                                 | 28.13                 | 1,407                              |
| 2.        | 110 – 150 | 8                                 | 25.00                 | 1,250                              |
| 3.        | 160 – 200 | 2                                 | 6.25                  | 313                                |
| 4.        | 210 – 250 | 2                                 | 6.25                  | 313                                |
| 5.        | 310 – 350 | 3                                 | 9.38                  | 469                                |
| 6.        | 410 – 450 | 2                                 | 6.25                  | 313                                |
| 7.        | 460 – 500 | 3                                 | 9.38                  | 469                                |
| 8.        | 510 – 550 | 1                                 | 3.13                  | 157                                |
| 9.        | 760 – 800 | 1                                 | 3.13                  | 157                                |
| 10.       | 860 – 900 | 1                                 | 3.13                  | 157                                |
| $\Sigma$  |           | <b>32</b>                         | <b>100.00</b>         | <b>5,005<sup>4</sup></b>           |

Anmerkungen.  $\rho = -.07$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten *k* für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt *k* von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt *k* = 5,005.

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .15 hat sich aus einem Verhältnis von 760 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,245 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Da die simulierten Effektgrößen keinerlei Publikationsbias aufweisen, kann hierbei ein Verhältnis von 15% signifikanten ( $p \leq .05$ , zweiseitig) zu 85% nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) als Referenz zur Beurteilung eines Publikationsbias herangezogen werden. Der aus den tatsächlich publizierten Effektgrößen berechnete Index beträgt .22 (sieben signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 25 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 25).



**Abbildung 25.**  $\rho = -.07$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 35) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie (McGrew, 2009).

Um bei den tatsächlich publizierten Effektgrößen einen Index von .15 – ohne Publikationsbias – zu erreichen, müssten 14 weitere nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von sieben signifikanten ( $p \leq .05$ , zweiseitig) zu 39 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) unter den gegebenen Parametern publiziert sein. Demzufolge kann in dem Fall ein Publikationsbias nicht ausgeschlossen werden.

*Moderatoranalysen:*

Zum Zusammenhang zwischen Gc und E beträgt die erklärte Varianz des artefaktkorrigierten Effektes ( $\rho$ ) ausnahmslos weniger als 75%. Dementsprechend wurden Moderatoranalysen durchgeführt. Die kognitiven Leistungstests berücksichtigend, wurden die Studien, in denen der GKT zur Messung von Gc verwendet wurde, mit den Studien verglichen, in denen andere kogLT als der GKT verwendet wurden (bspw. Watson-Glaser Critical Thinking Appraisal; Watson & Glaser, 1964). Darüber hinaus wurden Moderatoranalysen hinsichtlich der eingesetzten

## Kapitel 5 – Ergebnisse

Persönlichkeitsinventare durchgeführt. Dabei wurden folgende Persönlichkeitsinventare verglichen:

- FFM-Inventare vs.
- NEO-FFI vs.
- NEO-PI-(R) vs.
- Eysenck-Inventare.

In Tabelle 36 sind die metaanalytischen Kennwerte der entsprechenden Moderatoranalysen zum Zusammenhang zwischen Gc und E aufgeführt.

**Tabelle 36**  
Moderatoranalysen zum Zusammenhang zwischen Gc und E

| Gc/E              | Moderator  | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | %Var. | 90% CV | $\bar{r}$ | $SD_r$ | %Var  |
|-------------------|------------|----------|----------|--------|-------------|-------|--------|-----------|--------|-------|
| <i>kogLT</i>      |            |          |          |        |             |       |        |           |        |       |
| CHC <sup>1</sup>  | GKT        | 832      | 6        | -.02   | .035        | 84.56 | .03    | -.01      | .041   | 81.30 |
| CHC <sup>1</sup>  | ohne GKT   | 79,116   | 27       | .08    | .060        | 14.85 | .00    | .06       | .042   | 15.89 |
| CHC <sup>1a</sup> | ohne GKT   | 7,229    | 26       | -.08   | .071        | 51.30 | .01    | -.07      | .055   | 54.26 |
| <i>PSK</i>        |            |          |          |        |             |       |        |           |        |       |
| CHC <sup>1</sup>  | FFM        | 6,996    | 23       | -.06   | .068        | 48.55 | .03    | -.05      | .057   | 50.46 |
| CHC <sup>1</sup>  | NEO-PI-(R) | 3,110    | 11       | -.03   | .061        | 49.30 | .05    | -.03      | .061   | 49.04 |
| CHC <sup>1</sup>  | NEO-FFI    | 3,006    | 10       | -.11   | .069        | 54.94 | -.02   | -.08      | .056   | 51.04 |
| CHC <sup>1</sup>  | Eysenck    | 609      | 5        | -.18   | .046        | 86.25 | -.12   | -.14      | .032   | 88.83 |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und -überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; %Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde; <sup>1a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist; kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; GKT = General Knowledge Test; die Studien inkludiert, die den GKT zur Erfassung der kristallisierten Intelligenzausprägung verwendet haben; ohne GKT = die Studien inkludiert, die zur Erfassung der kristallisierten Intelligenz andere kogLT als den GKT verwendet haben; FFM = die Studien inkludiert, in denen „Big-Five“-Inventare eingesetzt wurden; NEO-PI-(R) = die Studien inkludiert, in denen der NEO-PI-R und der NEO-PI verwendet wurde; NEO-FFI = die Studien berücksichtigt, in denen der NEO-FFI eingesetzt wurde; Eysenck = die Studien inkludiert, die E mit Eysenck-Inventaren erfasst haben; grau markiert sind die Zeilen, in denen mehr als 75% der Varianz erklärt wird.

In den Studien, in denen der GKT zur Erfassung von Gc eingesetzt wurde, werden mehr als 75% der Varianz erklärt werden (Tabelle 36, grau markiert). Da die Null hierbei im Glaubwürdigkeitsintervall liegt, kann in dem Fall von einem Nulleffekt in dieser Subpopulation



ausgegangen werden (Tabelle 36). In den Studien, in denen andere kogLt als der GKT verwendet wurden, kann hingegen nicht ausgeschlossen werden, dass weitere Moderatoren existieren, die den Zusammenhang zwischen Gc und E erklären. Offenbar sind die konzeptionellen Unterschiede zwischen dem GKT und anderen kogLt in Bezug auf Gc relevant für die Zusammenhänge mit E. Hinsichtlich der Persönlichkeitsinventare werden in den Studien, in denen Eysenck-Inventare zur Erfassung der Extraversionsausprägung verwendet wurden, mehr als 75% der Varianz erklärt werden (Tabelle 36, grau markiert). Dabei kann von einem Zusammenhang von  $\rho = -.18$  für diese Subpopulation ausgegangen werden. Der Unterschied zwischen Eysenck-Inventaren und Nicht-Eysenck-Inventaren führt zu erheblicher Varianz der Effektstärken zum Zusammenhang zwischen Gc und E.

*Zusammenfassung zum Zusammenhang zwischen der Kristallisierten Intelligenz und der Extraversion:*

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = .09$ . Da die Null nicht im Glaubwürdigkeitsintervall liegt, kann von einem positiven Effekt in der Population ausgegangen werden, bei dem die Höhe aufgrund der erklärten Varianz von weniger als 75% nicht verallgemeinert werden sollte. Nach der anschließenden Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) verringert sich die artefaktkorrigierte Effektgröße auf  $\rho = -.07$ . Auch nach der Exklusion der Studie von Reeve et al. (2006) beträgt die erklärte Varianz weniger als 75%. Allerdings kann ein Populationseffekt von Null nicht mehr ausgeschlossen werden, das Glaubwürdigkeitsintervall die Null mit einschließt.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = .07$ . Nach der Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) verringert sich die korrigierte Effektgröße auf  $\rho = -.07$ . Unabhängig von der Exklusion oder Inklusion der Studie von Reeve et al. (2006) beträgt die erklärte Varianz weniger als 75% und die Null liegt im Glaubwürdigkeitsintervall. Dementsprechend kann davon ausgegangen werden, dass die artefaktkorrigierten Effektgrößen von  $\rho = .07$  bzw.  $\rho = -.07$  mittlere Effekte heterogener Subpopulationen sind, bei denen ein Nulleffekt in der Population nicht auszuschließen ist.

Hinsichtlich des Publikationsbias kann in beiden Analyse-Varianten nach der „trim- and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) eine symmetrische Verteilung der

publizierten Effektgrößen angenommen und demzufolge ein Publikationsbias ausgeschlossen werden. Nach den durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens, ist in beiden Analyse-Varianten ein Publikationsbias nicht auszuschließen, da zu viele signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) mit negativen Vorzeichen im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sind als man bei einer zufälligen Verteilung signifikanter ( $p \leq .05$ , zweiseitig) und nichtsignifikanter Effektgrößen ( $p > .05$ , zweiseitig) erwarten würde.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, konnten in den Studien, in denen der GKT oder Eysenck-Inventare verwendet wurden, jeweils mehr als 75% der Varianz erklärt werden. Es ist davon auszugehen, dass die konzeptionellen Unterschiede zwischen dem GKT und Nicht-GKT bzw. Eysenck-Inventaren und Nicht-Eysenck-Inventaren bezüglich des Zusammenhangs zwischen Gc und E relevant sind. Insbesondere führt der Unterschied zwischen Eysenck-Inventaren und Nicht-Eysenck-Inventaren zu erheblicher Varianz der Effektstärken.

### 5.7 Ergebnisse zur Allgemeinen Intelligenz und Offenheit für Erfahrung

*Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien:*

In dieser Analyse-Variante beträgt die inkludierte Studienanzahl  $k = 70$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{\text{bare-bones}} = .10$  ( $SD = .06$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 1,271$  (Range:  $n = 57$  bis 71,887;  $SD = 8,566.05$ ;  $N = 88,997$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. -überschätzung korrigierte Effektgröße beträgt  $\rho = .13$  ( $SD_{\rho} = .08$ ). Das Glaubwürdigkeitsintervall schließt die Null nicht ein und die erklärte Varianz beträgt 19.04%. (Tabelle 37).

Nach der anschließenden Exklusion der Studie von Reeve et al. (2006) mit einer Stichprobengröße von  $n = 71,887$  erhöht sich die unkorrigierte Effektgröße von  $\bar{r}_{\text{bare-bones}} = .10$  ( $SD = .06$ ) auf  $\bar{r}_{\text{bare-bones}} = .17$  ( $SD = .10$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 248$  (Range:  $n = 57$  bis 1,886;  $SD = 248.34$ ;  $N = 17,110$ ). Der Betrag der artefaktkorrigierten Effektgröße erhöht sich von  $\rho = .13$  ( $SD_{\rho} = .08$ ) auf  $\rho = .19$  ( $SD_{\rho} = .13$ ). Die Null liegt nicht im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 21.71% (Tabelle 37).

Unabhängig von der Inklusion oder Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) ist von einem positiven Zusammenhang zwischen g und O in der Population auszugehen. Allerdings

sollte die Höhe des positiven Zusammenhangs nicht generalisiert werden, da die erklärte Varianz deutlich weniger als 75% beträgt. Dementsprechend ist nicht auszuschließen, dass weitere Moderatoren existieren, die den Zusammenhang zwischen g und O erklären (siehe unten).

**Tabelle 37**

Metaanalytische Ergebnisse zum Zusammenhang zwischen g und O

| g/Offenheit                          | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | %Var. | 90% CV | $\bar{r}$ | $SD_r$ | %Var  |
|--------------------------------------|----------|----------|--------|-------------|-------|--------|-----------|--------|-------|
| kodiert nach Autor(en) <sup>1</sup>  | 88,997   | 70       | .13    | .075        | 19.04 | .03    | .10       | .057   | 19.35 |
| kodiert nach Autor(en) <sup>1a</sup> | 17,110   | 69       | .19    | .134        | 21.71 | .02    | .17       | .100   | 27.91 |
| kodiert nach CHC <sup>2</sup>        | 90,455   | 74       | .13    | .074        | 21.21 | .04    | .10       | .054   | 21.98 |
| kodiert nach CHC <sup>2a</sup>       | 18,565   | 73       | .20    | .132        | 25.74 | .03    | .16       | .092   | 30.49 |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und –überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; %Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist.

*Zuordnung der Fähigkeitskonstrukte gemäß CHC-Theorie:*

In dieser Analyse-Variante beträgt die inkludierte Studienzahl *k* = 74 bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .10$  ( $SD = .05$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 1,222$  (Range: *n* = 50 bis 71,887;  $SD = 8,329.13$ ; *N* = 90,455). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. –überschätzung korrigierte Effektgröße beträgt  $\rho = .13$  ( $SD_{\rho} = .07$ ). Die Null liegt nicht im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 21.21%.

Nach der Exklusion der Studie von Reeve et al. (2006, *n* = 71,887) verändert sich die unkorrigierte Effektgröße von  $\bar{r}_{bare-bones} = .10$  ( $SD = .06$ ) auf  $\bar{r}_{bare-bones} = .16$  ( $SD = .09$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 254$  (Range: *n* = 50 bis 900;  $SD = 183.83$ ; *N* = 18,565). Der Betrag der artefaktkorrigierten Effektgröße erhöhte sich durch die Exklusion der Studie von Reeve et al. (2006) von  $\rho = .13$  ( $SD_{\rho} = .07$ ) auf  $\rho = .20$  ( $SD_{\rho} = .13$ ). Die Null liegt hierbei ebenfalls nicht im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 25.74% (Tabelle 37).

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, kann ebenfalls – unabhängig von der Exklusion bzw.

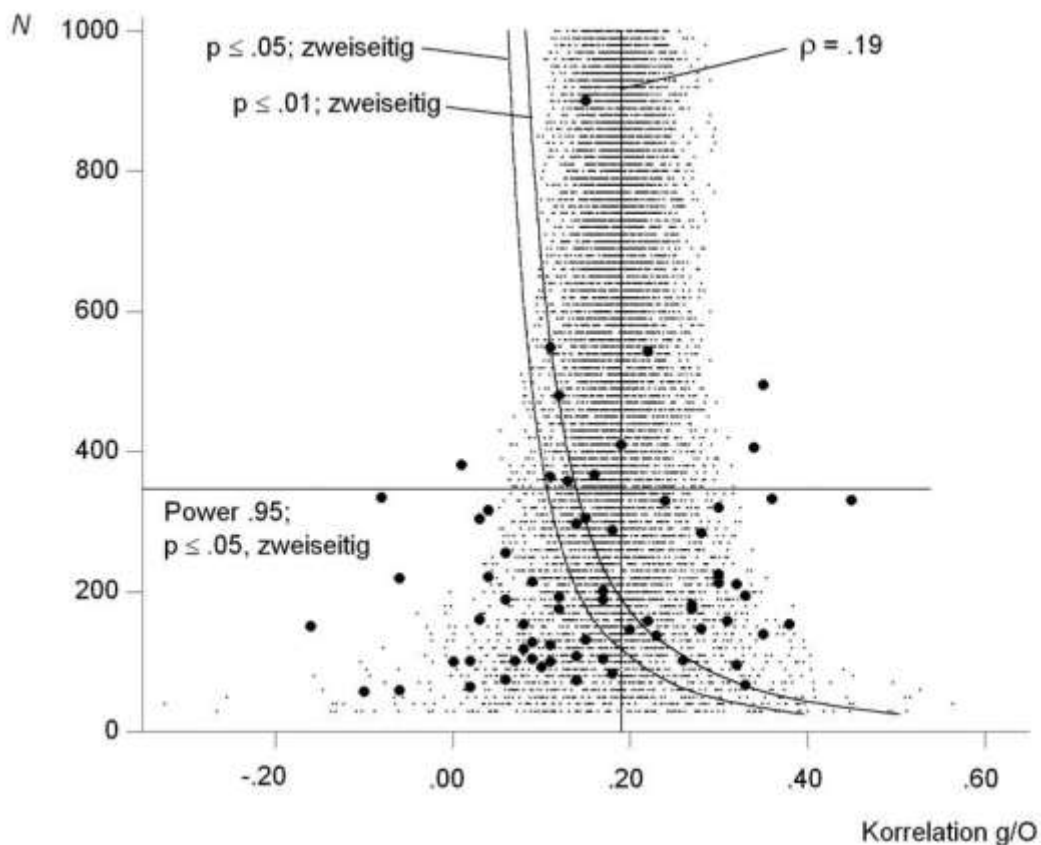
## *Kapitel 5 – Ergebnisse*

Inklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) – ein positiver Zusammenhang in der Population zwischen g und O angenommen werden. Ebenfalls kann in dieser Analyse-Variante nicht ausgeschlossen werden, dass weitere Moderatoren existieren, die den Zusammenhang zwischen g und O erklären. Dementsprechend wurden hier auch Moderatoranalysen durchgeführt (siehe unten).

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe der Schätzer mit  $R_0 < 1$  und  $L_0 < 1$  eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .16$  angenommen und demnach ein Publikationsbias ausgeschlossen werden.

Die Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens erlaubt keine schlüssige Interpretation der Ergebnisse hinsichtlich eines Publikationsbias. Dementsprechend wurde, wie bereits zwischen g und Ne (Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie vorgenommen wurde) auf die Simulationsmethode zur Identifikation des Publikationsbias verzichtet.



**Abbildung 26.**  $\rho = .19$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet,  $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4; GPower (Faul & Erdfelder, 1992; Power .95,  $p \leq .05$ , zweiseitig):  $n = 349$ ; eingezeichnete Signifikanzfunktionen mit  $p \leq .05$  (zweiseitig) und  $p \leq .01$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien.

Anhand der Abbildung 26 ist auffällig, dass vier der tatsächlich publizierten Effektgrößen im linken und vier der Effektgrößen im rechten Bereich, deutlich außerhalb der simulierten Effektgrößen, liegen. In Tabelle 38 sind Angaben zu den Studien aufgeführt, in denen Effektgrößen berichtet werden, die außerhalb der simulierten Effektgrößen liegen. Allerdings lassen sich keine eindeutigen Auffälligkeiten bzw. Gemeinsamkeiten feststellen, weshalb diese Studien Effektgrößen berichten, die außerhalb der simulierten Effektgrößen liegen.

**Tabelle 38**

Deskriptive Angaben publizierter Studien, in denen die Effektgröße des Zusammenhangs zwischen g und O von den gezogenen Effektgrößen aus der adaptierten Simulation abweicht

| Studie   | kogLT             | PSK          | Land    | Alter  | Stichprobe                           | Thema  |
|--|-------------------|--------------|---------|--|--------------------------------------|--|
| Allik und Realo (1997) <sup>1</sup>                      | unbekannter kogLT | NEO-PI       | Estland | $M = 19.10$<br>( $SD = 2.30$ )                     | angehende Studenten<br>( $n = 381$ ) | akademische Fähigkeiten  |
| LePine (2003) <sup>1</sup>                               | WPT               | NEO-PI-R     | USA     | $M = 21.00$<br>( $SD = 2.29$ )                     | Studenten<br>( $n = 219$ )           | Teamadaptation und -performanz   |
| Furnham und Monsen (2009) <sup>1</sup>                   | WPT               | NEO-PI-R     | GB      | $M = n.b.$<br>( $SD = n.b.$ )<br>Range:<br>15 - 16 | Schüler<br>( $n = 334$ )             | Vorhersage von Schulnoten durch Persönlichkeit und Intelligenz             |
| Ziegler, Schmukle, Egloff und Bühner (2010) <sup>1</sup> | IST-2000-R        | NEO-PI-R     | D/A     | $M = 22.32$<br>( $SD = 5.56$ )                     | Studenten<br>( $n = 150$ )           | unterschiedliche Messungen der Leistungsmotivation                         |
| Maurer, Lippstreu und Judge (2008) <sup>2</sup>          | VT + DR           | IPIP         | USA     | $M = 39.32$<br>( $SD=11.14$ )                      | Angestellte<br>( $n = 330$ )         | Fertigkeiten in Abhängigkeit der Beteiligung und individueller Differenzen |
| Austin et al. (2002) <sup>2,3</sup>                      | Testbatterie      | short NEO-PI | SE      | $M > 50.00$<br>( $SD = n.b.$ )                     | Zwillinge<br>( $n = 495$ )           | Intelligenz und persönliche/soziale Adjustierung                           |
| Austin et al. (2002) <sup>2,4</sup>                      | Testbatterie      | NEO-FFI      | GB      | $M = n.b.$<br>( $SD = n.b.$ )<br>Range:<br>55 - 74 | keine Angabe<br>( $n = 405$ )        | Intelligenz und persönliche/soziale Adjustierung                           |
| Ackerman, Bowen, Beier und Kanfer (2001) <sup>2</sup>    | Testbatterie      | NEO-FFI      | USA     | $M = 19.00$<br>( $SD = 0.41.$ )                    | Studenten<br>( $n = 320$ )           | Vorhersage von Intelligenz durch Persönlichkeit                            |

*Anmerkungen.* kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; TTA = Thurstone Test of Alertness; VT+DR = Autoren haben aus einen „Vocabulary“-Test und „Deductive-Reasoning“-Test einen g-Faktor gebildet; <sup>1</sup> = publizierte Effektgrößen, die außerhalb der simulierten Effektgrößen im linken Bereich liegen (siehe Abbildung 26); <sup>2</sup> = publizierte Effektgrößen, die außerhalb der simulierten Effektgrößen im rechten Bereich liegen (siehe Abbildung 26); <sup>3</sup> = Datensatz 1 der Studie von Austin et al. (2002); <sup>4</sup> = Datensatz 2 der Studie von Austin et al. (2002).

In der Abbildung 26 wurde zusätzlich die Signifikanzfunktion mit  $p \leq .01$  (zweiseitig) eingezeichnet. Dabei ist auffällig, dass ca. sieben publizierte Effektgrößen auf der eingezeichneten Signifikanzfunktion ( $p \leq .01$ , zweiseitig) liegen. Weiterhin ist auffällig, dass sich ca. 13 publizierte signifikante Effektgrößen mit  $p \leq .01$  (zweiseitig) und/oder  $p \leq .05$  (zweiseitig) entlang der eingezeichneten Signifikanzfunktionen ( $p \leq .05$ , zweiseitig;  $p \leq .01$ , zweiseitig) verteilen (Abbildung 26). Diese Verteilung entlang der Signifikanzfunktion tritt allerdings nicht im Bereich von  $p > .05$  (zweiseitig) auf. Möglicherweise besteht ein Bias im nichtsignifikanten Wertebereich ( $p > .05$ , zweiseitig), entlang der eingezeichneten Signifikanzfunktion ( $p \leq .05$ , zweiseitig; Abbildung 26), was auf ein Bemühen hindeuten kann,

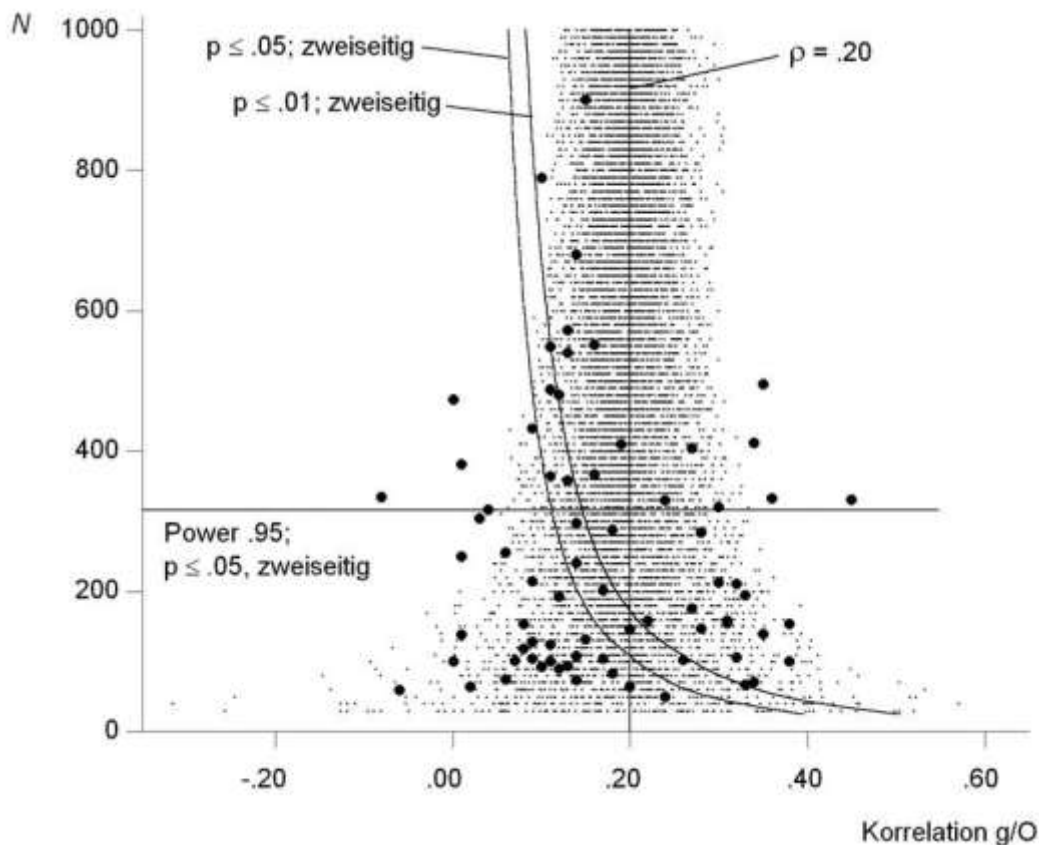
## *Kapitel 5 – Ergebnisse*

dass Effektgrößen, die zunächst in der Nähe der Signifikanzgrenze auftreten, im nachhinein durch Vergrößerung der Stichprobe in den signifikanten Bereich „gedrückt“ werden (vgl. Furguson, 2009; Kirk, 1996).

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durchgeführt wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) kann auf Grund der Höhe des Schätzer mit  $R_0 < 1$  von einer symmetrischen Verteilung um die mittlere Effektgröße von  $\bar{r} = .17$  ausgegangen und demnach ein Publikationsbias ausgeschlossen werden. Das zweite Symmetriemaß zeigt allerdings einen Wert von  $L_0 = 1.70$  um die mittlere Effektgröße von  $\bar{r} = .17$  an. Demnach wäre ein Publikationsbias nicht auszuschließen. Nach der anschließenden Exklusion von zwei publizierten Effektgrößen ( $L_0^+ = 2$ ) mit negativen Vorzeichen und den höchsten Beträgen kann von einer symmetrischen Verteilung der Effektgrößen hinsichtlich der Summe aller Ränge (siehe Anhang A4) um die mittlere Effektgröße von  $\bar{r} = .18$  mit  $L_0 < 1$  ausgegangen werden. Ein Publikationsbias wäre demnach auszuschließen.

Die Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens erlaubt keine schlüssige Interpretation der Ergebnisse hinsichtlich eines Publikationsbias zum Zusammenhang zwischen  $g$  und  $O$ . Dementsprechend wurde hierbei auf die Ergebnisdarstellung der durchgeführten Simulation zur Identifikation eines Publikationsbias verzichtet.



**Abbildung 27.**  $\rho = .20$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet,  $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4; eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig) und  $p \leq .01$  (zweiseitig); GPower (Faul & Erdfelder, 1992; Power .95,  $p \leq .05$ , zweiseitig):  $n = 314$ ; Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie (McGrew, 2009).

Anhand der Abbildung 27 ist auffällig, dass sechs der tatsächlich publizierten Effektgrößen im linken und vier der Effektgrößen im rechten Bereich, außerhalb der simulierten Effektgrößen, liegen. Deskriptive Angaben zu den Studien, die Effektgrößen publiziert haben, die hinsichtlich ihrer Höhe außerhalb der simulierten Effektgrößen liegen, sind in Tabelle 38 (siehe oben) aufgeführt. Darüber hinaus sind in Tabelle 39 Angaben zu den Studien aufgeführt, in denen Effektgrößen berichtet werden, die im linken Bereich, außerhalb der simulierten Effektgrößen liegen. Auffälligkeiten bzw. Gemeinsamkeiten, weshalb die Studien (Tabelle 39) Effektgrößen berichten, die außerhalb der simulierten Effektgrößen liegen, konnten allerdings auch durch Sichtung der entsprechenden Studien nicht festgestellt werden.



**Tabelle 39**

Deskriptive Angaben publizierter Studien, in denen die Effektgröße des Zusammenhangs zwischen g und O von den gezogenen Effektgrößen aus der adaptierten Simulation abweicht

| Studie                               | kogLT                | PSK      | Land    | Alter  | Stichprobe                               | Thema   |
|--------------------------------------|----------------------|----------|---------|--|--|---|
| Kickul und Neumann (2000)            | TTA                  | IPIP     | USA     | $M = n.b.$<br>( $SD = n.b.$ )                      | Studenten<br>( $n = 304$ )               | Führungsverhalten:<br>Persönlichkeit und<br>Intelligenz                     |
| Neuman und Wright (1999)             | TTA                  | NEO-PI-R | USA     | nicht<br>spezifiziert                              | Personal-<br>Management<br>( $n = 316$ ) | Team Effektivität   |
| Allik und Realo (1997)               | unbekannter<br>kogLT | NEO-PI   | Estland | $M = 19.10$<br>( $SD = 2.30$ )                     | angehende<br>Studenten<br>( $n = 381$ )  | akademische<br>Fähigkeiten  |
| Pearman (2009)                       | Testbatterie         | DA       | USA     | $M = 47.80$<br>( $SD=13.21$ )                      | n.b.<br>( $n = 250$ )                    | Effekte des<br>Geschlechts und der<br>Persönlichkeit im<br>Erwachsenenalter |
| Furnham und Monsen (2009)            | WPT                  | NEO-PI-R | GB      | $M = n.b.$<br>( $SD = n.b.$ )<br>Range:<br>15 - 16 | Schüler<br>( $n = 334$ )                 | Vorhersage von<br>Schulnoten durch<br>Persönlichkeit und<br>Intelligenz     |
| Chamorro-Premuzic und Arteche (2008) | WPT+BRT              | NEO-PI-R | GB      | $M = 20.31$<br>( $SD = 3.34$ )                     | Studenten<br>( $n = 473$ )               | intellektuelle<br>Kompetenzen und<br>akademische<br>Performanz              |

*Anmerkungen.* kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; TTA = Thurstone Test of Alertness; WPT+BRT = Autor der vorliegenden Arbeit hat aus den Wondelic Personel Test und dem Baddeley Reasoning Test g gebildet; DA = „Descriptives Attributes“-FFM-Inventar).

In Abbildung 27 ist ebenfalls wie in Abbildung 26 eine weitere Signifikanzfunktion mit  $p \leq .01$  (zweiseitig) eingezeichnet. Dabei kann es auffallen, dass sich ca. 14 publizierte signifikante Effektgrößen mit  $p \leq .01$  (zweiseitig) und/oder  $p \leq .05$  (zweiseitig) entlang der eingezeichneten Signifikanzfunktionen ( $p \leq .05$ , zweiseitig;  $p \leq .01$ , zweiseitig) verteilen. Auch in dieser Analyse-Variante kann der Effekt, dass ein Bemühen bestehen könnte, Effektgrößen, die zunächst in der Nähe der Signifikanzgrenze auftreten, im nachhinein durch Vergrößerung der Stichprobe signifikant zu bekommen, nicht ausgeschlossen werden.

*Moderatoranalysen:*

Zum Zusammenhang zwischen g und O beträgt die Varianz des artefaktkorrigierten Effektes ( $\rho$ ) ausnahmslos weniger als 75%. Die kognitiven Leistungstests der publizierten Studien berücksichtigend, wurde ein Moderator gerechnet, bei dem ausschließlich die Studien inkludiert wurden, in denen der WPT zur Erfassung der allgemeinen Intelligenzausprägung eingesetzt wurde. Die metaanalytischen Kennwerte dieses Moderators wurden mit den Kennwerten verglichen, bei denen zur Messung der allgemeinen Intelligenzausprägung andere kogLT als der WPT eingesetzt wurde. Die entsprechenden Ergebnisse sind in Tabelle 40 aufgeführt. Darüber hinaus wurden

## *Kapitel 5 – Ergebnisse*

Moderatoranalysen hinsichtlich der eingesetzten Persönlichkeitsinventare durchgeführt. Dabei wurden folgende Persönlichkeitsinventare verglichen:

- NEO-Inventare vs.
- NEO-PI-(R) vs.
- NEO-FFI vs.
- FFM-Inventare, allerdings keine NEO-Inventare.

In Tabelle 40 sind die metaanalytischen Kennwerte der entsprechenden Moderatoranalysen zum Zusammenhang zwischen g und O aufgeführt.

**Tabelle 40**

Moderatoranalysen zum Zusammenhang zwischen g und O

| g/O                     | Moderator   | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | %Var. | 90% CV | $\bar{r}$ | $SD_r$ | %Var  |
|-------------------------|-------------|----------|----------|--------|-------------|-------|--------|-----------|--------|-------|
| <u>kogLT</u>            |             |          |          |        |             |       |        |           |        |       |
| Autor(en) <sup>1</sup>  | WPT         | 4,895    | 28       | .16    | .093        | 50.94 | .04    | .13       | .071   | 52.53 |
| Autor(en) <sup>1</sup>  | ohne WPT    | 84,102   | 42       | .13    | .074        | 13.48 | .03    | .10       | .055   | 13.80 |
| Autor(en) <sup>1a</sup> | ohne WPT    | 12,215   | 41       | .20    | .144        | 15.57 | .01    | .19       | .103   | 22.77 |
| CHC <sup>2</sup>        | WPT         | 4,676    | 27       | .17    | .086        | 53.06 | .05    | .14       | .061   | 59.87 |
| CHC <sup>2</sup>        | ohne WPT    | 85,689   | 47       | .13    | .072        | 15.91 | .04    | .10       | .052   | 16.54 |
| CHC <sup>2a</sup>       | ohne WPT    | 13,802   | 46       | .21    | .143        | 20.19 | .03    | .17       | .100   | 24.19 |
| <u>PSK</u>              |             |          |          |        |             |       |        |           |        |       |
| Autor(en) <sup>1</sup>  | FFM, NEO    | 11,655   | 46       | .18    | .121        | 23.96 | .02    | .17       | .097   | 28.42 |
| Autor(en) <sup>1</sup>  | NEO-PI-(R)  | 7,874    | 27       | .16    | .110        | 23.88 | .02    | .15       | .102   | 23.89 |
| Autor(en) <sup>1</sup>  | NEO-FFI     | 3,781    | 19       | .22    | .140        | 26.20 | .04    | .20       | .077   | 44.05 |
| Autor(en) <sup>1</sup>  | FFM, no NEO | 4,929    | 21       | .20    | .143        | 22.51 | .01    | .17       | .093   | 31.54 |
| CHC <sup>2</sup>        | FFM, NEO    | 13,448   | 52       | .19    | .116        | 29.74 | .04    | .16       | .086   | 33.11 |
| CHC <sup>2</sup>        | NEO-PI-(R)  | 7,523    | 31       | .18    | .116        | 28.96 | .03    | .15       | .098   | 29.34 |
| CHC <sup>2</sup>        | NEO-FFI     | 5,925    | 21       | .21    | .112        | 32.28 | .07    | .17       | .069   | 41.63 |
| CHC <sup>2</sup>        | FFM, no NEO | 4,504    | 19       | .21    | .151        | 24.11 | .02    | .16       | .095   | 30.75 |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und –überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; %Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von g von den Autoren der publizierten Studien vorgenommen wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist; kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; WPT = Wonderlic Personel Test (hierbei die Studien inkludiert, die den WPT zur Erfassung der allgemeinen Intelligenzausprägung verwendet haben; ohne WPT = die Studien inkludiert, die zur Erfassung der allgemeinen Intelligenz andere kogLT als den WPT verwendet haben; FFM, NEO = die Studien inkludiert, in denen Costa und McCrae-Inventare (NEO-PI, NEO-PI-R, NEO-FFI) eingesetzt wurden; NEO-PI-(R) = die Studien inkludiert, in denen der NEO-PI-R und der NEO-PI verwendet wurde; NEO-FFI = die Studien berücksichtigt, in denen der NEO-FFI eingesetzt wurde; FFM, no NEO = die Studien berücksichtigt, in denen FFM-Inventare, allerdings keine Costa und McCrae Inventare verwendet wurden.

Mit keinem, im Rahmen der vorliegenden Arbeit, gerechneten Moderator konnten mehr als 50.94% der Varianz erklärt werden (Tabelle 40). Offenbar existieren weitere Moderatoren, die den Zusammenhang zwischen g und O erklären.

## Kapitel 5 – Ergebnisse

### *Zusammenfassung zum Zusammenhang zwischen der Allgemeinen Intelligenz und Offenheit für Erfahrung:*

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien beibehalten wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = .13$ . Nach der Exklusion der Studie von Reeve et al. (2006) mit einer Stichprobengröße von  $n = 71,887$  beträgt die korrigierte Effektgröße  $\rho = .19$ . Unabhängig von der Exklusion oder Inklusion der Studie von Reeve et al. (2006) liegt die Null nicht im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt weniger als 75%. Dementsprechend kann von in der Population von einem positiven Zusammenhang zwischen g und O ausgegangen werden, allerdings sollte die Höhe des Populationszusammenhangs nicht generalisiert werden.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, beträgt die korrigierte Effektgröße  $\rho = .13$ . Nach der Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) beträgt die korrigierte Effektgröße  $\rho = .20$ . Auch in dieser Analyse-Variante liegt die Null nicht im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt weniger als 75%. Dementsprechend ist von einem positiven Populationszusammenhang auszugehen, bei dem die Höhe nicht zu generalisieren ist.

Hinsichtlich des Publikationsbias kann nach der „trim- and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde, eine symmetrische Verteilung der publizierten Effektgröße um die mittlere Effektgröße von  $\bar{r} = .16$  angenommen und dementsprechend ein Publikationsbias ausgeschlossen werden. In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, kann nach der „trim- and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) aufgrund des Schätzers  $R_0 < 1$  eine symmetrische Verteilung der publizierten Effektgröße um die mittlere Effektgröße von  $\bar{r} = .17$  angenommen und ein Publikationsbias ausgeschlossen werden. Aufgrund des zweiten Schätzers mit  $L_0 > 1$  sollte hingegen ein Publikationsbias nicht ausgeschlossen werden.

Mit keinem Moderator wurden mehr als 75% der Varianz erklärt. Offenbar existieren weitere, in der vorliegenden Arbeit nicht identifizierte Moderatoren, die den Zusammenhang zwischen g und O erklären.

### 5.8 Ergebnisse zur Fluiden Intelligenz und Offenheit für Erfahrung

#### *Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien:*

In dieser Analyse-Variante beträgt die inkludierte Studienanzahl  $k = 28$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{\text{bare-bones}} = .04$  ( $SD = .04$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 1,001$  (Range:  $n = 64$  bis  $20,175$ ;  $SD = 3,791.56$ ;  $N = 28,035$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. -überschätzung korrigierte Effektgröße beträgt  $\rho = .06$  ( $SD_{\rho} = .05$ ). Das Glaubwürdigkeitsintervall schließt die Null mit ein und die erklärte Varianz beträgt 42.66%. (Tabelle 41).

Nach der anschließenden Exklusion der Studie von De Fruyt et al. (2006) mit einer Stichprobengröße von  $n = 20,175$  erhöht sich die unkorrigierte Effektgröße von  $\bar{r}_{\text{bare-bones}} = .04$  ( $SD = .04$ ) auf  $\bar{r}_{\text{bare-bones}} = .09$  ( $SD = .05$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 291$  (Range:  $n = 64$  bis  $2,658$ ;  $SD = 503.11$ ;  $N = 7,860$ ). Der Betrag der artefaktkorrigierten Effektgröße erhöht sich hierbei von  $\rho = .06$  ( $SD_{\rho} = .05$ ) auf  $\rho = .09$  ( $SD_{\rho} = .06$ ). Die Null liegt nicht im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 50.79% (Tabelle 41).

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien berücksichtigt wurde, kann ein Nulleffekt in der Population nicht ausgeschlossen werden. Nach der Exklusion einer einzigen Studie (De Fruyt et al., 2006,  $n = 20,175$ ) kann hingegen von einem positiven Effekt in der Population ausgegangen werden. Unabhängig von der Exklusion bzw. Inklusion der Studie von De Fruyt et al. (2006) kann aufgrund der erklärten Varianz, die kleiner als 75% ist, nicht ausgeschlossen werden, dass weitere Moderatoren existieren, die den Effekt zum Zusammenhang zwischen Gf und O erklären. Dementsprechend wurden Moderatoranalysen durchgeführt (siehe unten).

**Tabelle 41**

Metaanalytische Ergebnisse zum Zusammenhang zwischen Gf und O

| Gf/Offenheit                         | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | %Var. | 90% CV | $\bar{r}$ | $SD_r$ | %Var  |
|--------------------------------------|----------|----------|--------|-------------|-------|--------|-----------|--------|-------|
| kodiert nach Autor(en) <sup>1</sup>  | 28,035   | 28       | .06    | .050        | 42.66 | .00    | .04       | .042   | 36.01 |
| kodiert nach Autor(en) <sup>1a</sup> | 7,860    | 27       | .09    | .058        | 50.79 | .02    | .09       | .053   | 54.96 |
| kodiert nach CHC <sup>2</sup>        | 39,499   | 51       | .08    | .061        | 37.26 | .00    | .06       | .052   | 32.39 |
| kodiert nach CHC <sup>2a</sup>       | 19,324   | 50       | .10    | .062        | 45.04 | .02    | .09       | .050   | 50.13 |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und -überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; %Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße; <sup>1</sup> = Datensätze, in denen eine Zuordnung von Gf explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von Gf explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von De Fruyt et al. (2006, *n* = 20,175) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde und die Studie von De Fruyt et al. (2006, *n* = 20,175) exkludiert ist.

*Zuordnung der Fähigkeitskonstrukte gemäß CHC-Theorie:*

In dieser Analyse-Variante beträgt die inkludierte Studienzahl *k* = 51 bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .06$  ( $SD = .05$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 775$  (Range: *n* = 64 bis 20,175;  $SD = 2,832.64$ ; *N* = 39,499). Die artefaktkorrigierte Effektgröße beträgt  $\rho = .08$  ( $SD_{\rho} = .06$ ). Das Glaubwürdigkeitsintervall schließt die Null mit ein und die erklärte Varianz beträgt 37.26%.

Nach der anschließenden Exklusion der Studie von De Fruyt et al. (2006, *n* = 20,175), erhöht sich die unkorrigierte Effektgröße von  $\bar{r}_{bare-bones} = .06$  ( $SD = .05$ ) auf  $\bar{r}_{bare-bones} = .09$  ( $SD = .05$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 386$  (Range: *n* = 64 bis 2,982;  $SD = 587.11$ ; *N* = 19,324). Der Betrag der artefaktkorrigierten Effektgröße erhöhte sich durch die Exklusion der Studie von De Fruyt et al. (2006) von  $\rho = .08$  ( $SD_{\rho} = .06$ ) auf  $\rho = .10$  ( $SD_{\rho} = .06$ ). Die Null liegt hierbei nicht im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 45.04% (Tabelle 41).

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, kann ein Nulleffekt in der Population nicht ausgeschlossen werden. Nach der Exklusion der Studie von De Fruyt et al. (2006, *n* = 20,175) kann hingegen von einem positiven Populationszusammenhang ausgegangen werden. Auch in dieser Analyse-Variante beträgt die erklärte Varianz weniger als 75%. Dementsprechend wurden Moderatoranalysen durchgeführt (siehe unten).

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe des Schätzer mit  $R_0 = 1$  von einer asymmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .12$  ausgegangen und ein Publikationsbias nicht ausgeschlossen werden. Nach der anschließenden Exklusion einer Effektgröße mit positiven Vorzeichen und den höchsten Betrag kann eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere, neu bestimmte Effektgröße von  $\bar{r} = .11$  mit  $R_0 < 1$  angenommen und ein Publikationsbias ausgeschlossen werden. Der zweite Schätzer zeigt mit  $L_0 = 3.55$  ebenfalls eine asymmetrische Verteilung um die mittlere Effektgröße von  $\bar{r} = .12$  an. Nach der Exklusion von vier Effektgrößen, allerdings mit negativen Vorzeichen und den höchsten Beträgen, kann eine symmetrische Verteilung um die mittlere, neu bestimmte Effektgröße von  $\bar{r} = .14$  angenommen und ein Publikationsbias ausgeschlossen werden. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 26 Stichprobengrößen real publizierter Studien und die entsprechende Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 42 aufgeführt.

**Tabelle 42**

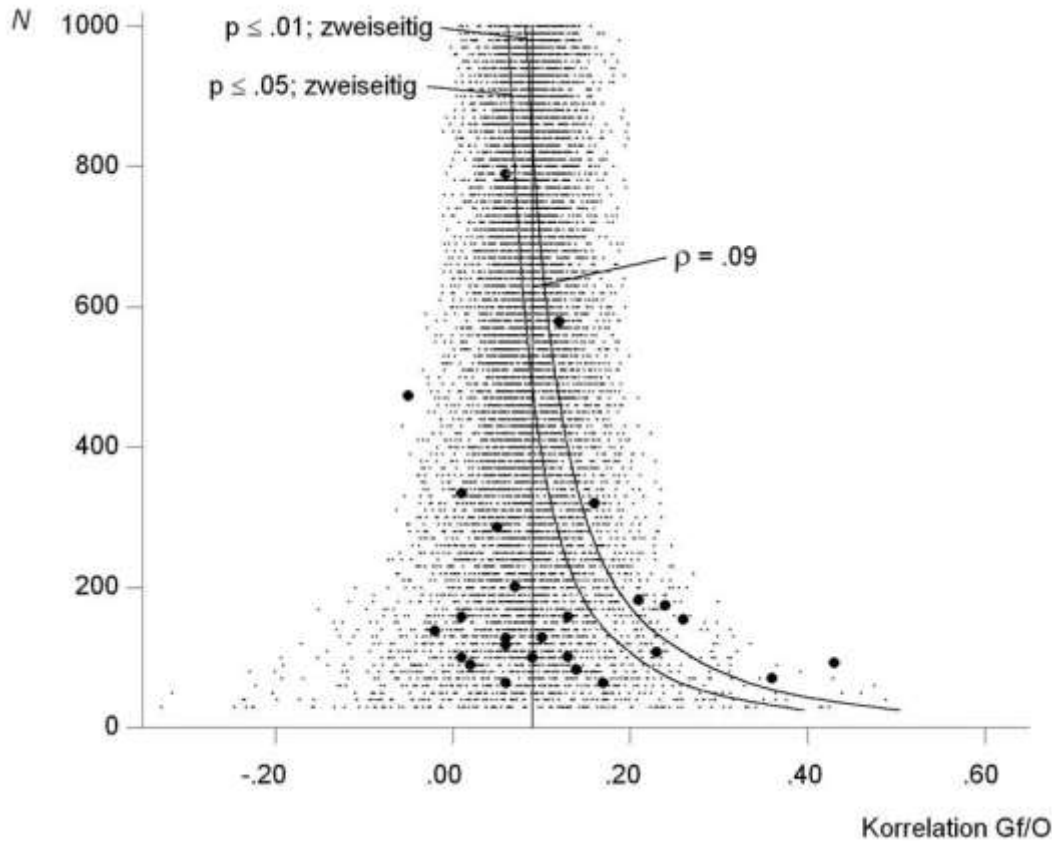
Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | <i>N</i>  | <i>k</i> (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | <i>k</i> (Simulation) <sup>3</sup> |
|-----------|-----------|-----------------------------------|-----------------------|------------------------------------|
| 1.        | 60 – 100  | 9                                 | 34.62                 | 1,731                              |
| 2.        | 110 – 150 | 6                                 | 23.08                 | 1,154                              |
| 3.        | 160 – 200 | 5                                 | 19.23                 | 962                                |
| 4.        | 260 – 300 | 1                                 | 3.85                  | 193                                |
| 5.        | 310 – 350 | 2                                 | 7.69                  | 385                                |
| 6.        | 460 – 500 | 1                                 | 3.85                  | 193                                |
| 7.        | 560 – 600 | 1                                 | 3.85                  | 193                                |
| 8.        | 760 – 800 | 1                                 | 3.85                  | 193                                |
| $\Sigma$  |           | <b>26</b>                         | <b>100.00</b>         | <b>5,004<sup>4</sup></b>           |

*Anmerkungen.*  $p = .09$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten *k* für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt *k* von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt *k* = 5,004.

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .18 hat sich aus einem Verhältnis von 925 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,079 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Da die simulierten Effektgrößen keinerlei Publikationsbias aufweisen, kann hierbei ein Verhältnis von 18% signifikanten ( $p \leq .05$ , zweiseitig) zu 82% nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) als Referenz zur

Beurteilung eines Publikationsbias herangezogen werden. Der aus den tatsächlich publizierten Effektgrößen berechnete Index beträgt .31 (acht signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 18 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 28).



**Abbildung 28.**  $\rho = .09$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 42) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig) und  $p \leq .01$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien.

Um bei den tatsächlich publizierten Effektgrößen einen Index von .18 – ohne Publikationsbias – zu erreichen, müssten 19 weitere nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von acht signifikanten ( $p \leq .05$ , zweiseitig) zu 37 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) unter den gegebenen Parametern publiziert sein. Ein Publikationsbias sollte in dem Fall nicht ausgeschlossen werden. Es sind zu viele Studien mit signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde.



In Abbildung 28 wurde neben der Signifikanzfunktion mit  $p \leq .05$  (zweiseitig) eine weitere Signifikanzfunktion mit  $p \leq .01$  (zweiseitig) eingezeichnet. Dabei kann es auffallen, dass sich sieben Effektgrößen entlang der eingezeichneten Signifikanzfunktionen mit  $p \leq .01$  (zweiseitig) verteilen und eine signifikante Effektgröße ( $p \leq .05$ , zweiseitig) nahe der eingezeichneten Signifikanzfunktionen mit  $p \leq .05$  (zweiseitig) liegt. Gleichzeitig verteilen sich sind nur zwei nichtsignifikante Effektgröße ( $p > .05$ , zweiseitig) entlang der eingezeichneten Signifikanzfunktionen mit  $p \leq .05$  (zweiseitig). Deshalb kann ein Bias im nichtsignifikanten Wertebereich ( $p > .05$ , zweiseitig), entlang der eingezeichneten Signifikanzfunktion ( $p \leq .05$ , zweiseitig), nicht ausgeschlossen werden. Möglicherweise zeigt sich darin ein Bemühen, durch eine Vergrößerung der Stichprobe, nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in den signifikanten Bereich zu „drücken“ (vgl. Furguson, 2009; Abbildung 28).

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durchgeführt wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe des Schätzer mit  $R_0 < 1$  eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .12$  angenommen und ein Publikationsbias ausgeschlossen werden. Der zweite Schätzer zeigt hingegen mit  $L_0 = 1.80$  eine asymmetrische Verteilung an. Dementsprechend ist ein Publikationsbias nicht auszuschließen. Durch die Exklusion von zwei Effektgrößen ( $L_0^+ = 2$ ) mit negativen Vorzeichen und den größten Beträgen kann eine Symmetrie der summierten Ränge der publizierten Effektgrößen um die neu bestimmte, jedoch unveränderte Effektgröße von  $\bar{r} = .12$  angenommen und ein Publikationsbias ausgeschlossen werden. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 47 relevanten Stichprobengrößen real publizierter Studien und die entsprechende Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 43 aufgeführt.

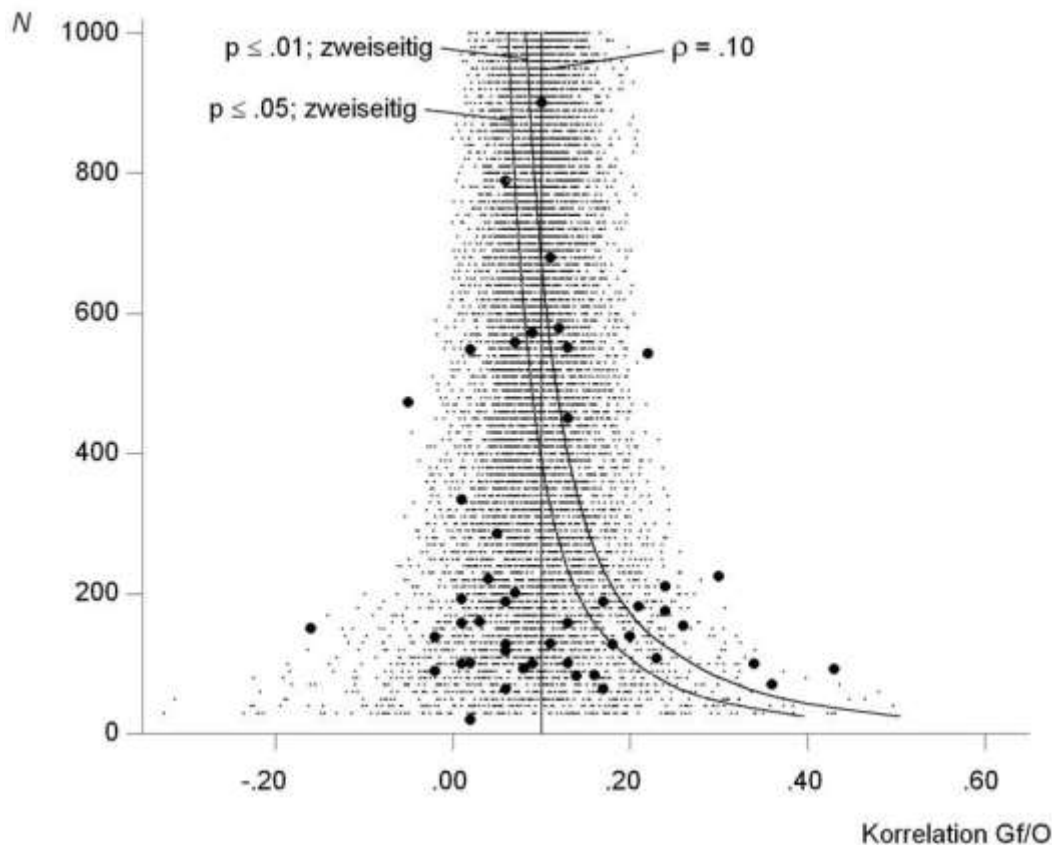
**Tabelle 43**

Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | <i>N</i>  | <i>k</i> (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | <i>k</i> (Simulation) <sup>3</sup> |
|-----------|-----------|-----------------------------------|-----------------------|------------------------------------|
| 1.        | 60 – 100  | 13                                | 27.66                 | 1,383                              |
| 2.        | 110 – 150 | 9                                 | 19.12                 | 956                                |
| 3.        | 160 – 200 | 9                                 | 19.12                 | 956                                |
| 4.        | 210 – 250 | 3                                 | 6.38                  | 319                                |
| 5.        | 260 – 300 | 1                                 | 2.13                  | 107                                |
| 6.        | 310 – 350 | 1                                 | 2.13                  | 107                                |
| 7.        | 410 – 450 | 1                                 | 2.13                  | 107                                |
| 8.        | 460 – 500 | 1                                 | 2.13                  | 107                                |
| 9.        | 510 – 550 | 3                                 | 6.38                  | 319                                |
| 10.       | 560 – 600 | 3                                 | 6.38                  | 319                                |
| 11.       | 660 – 700 | 1                                 | 2.13                  | 107                                |
| 12.       | 760 – 800 | 1                                 | 2.13                  | 107                                |
| 13.       | 860 – 900 | 1                                 | 2.13                  | 107                                |
| <b>Σ</b>  |           | <b>47</b>                         | <b>100.00</b>         | <b>5,001<sup>4</sup></b>           |

Anmerkungen.  $p = .10$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten *k* für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt *k* von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt *k* = 5,001.

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .28 hat sich aus einem Verhältnis von 1,394 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 3,607 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Da die simulierten Effektgrößen keinerlei Publikationsbias aufweisen, kann hierbei ein Verhältnis von 28% signifikanten ( $p \leq .05$ , zweiseitig) zu 72% nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) als Referenz zur Beurteilung eines Publikationsbias herangezogen werden. Der aus den tatsächlich publizierten Effektgrößen berechnete Index beträgt .40 (19 signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 28 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 29).



**Abbildung 29.**  $\rho = .10$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 43) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig) und  $p \leq .01$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie (McGrew, 2009).

Um bei den tatsächlich publizierten Effektgrößen einen Index von .28 – ohne Publikationsbias – zu erreichen, müssten 20 weitere nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von 19 signifikanten ( $p \leq .05$ , zweiseitig) zu 48 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) unter den gegebenen Parametern publiziert sein. Ein Publikationsbias kann auch in dieser Analyse-Variante nicht ausgeschlossen werden, da unter den gegebenen Parametern zu viele Studien mit signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt sind.

In Abbildung 29 wurde neben der Signifikanzfunktion mit  $p \leq .05$  (zweiseitig) eine weitere Signifikanzfunktion mit  $p \leq .01$  (zweiseitig) eingezeichnet. Dabei ist auffällig, dass sich ca. sieben signifikante Effektgrößen mit  $p \leq .01$  (zweiseitig) nahe der Signifikanzfunktion ( $p \leq .01$ , zweiseitig) verteilen. Darüber hinaus verteilen sich fünf signifikante Effektgrößen mit  $p \leq .05$  (zweiseitig) der Signifikanzfunktion ( $p \leq .05$ , zweiseitig, Abbildung 29). Hingegen verteilen sich lediglich ca. zwei bis drei nichtsignifikante Effektgrößen mit  $p > .05$  (zweiseitig) nahe der

## *Kapitel 5 – Ergebnisse*

eingezeichneten Signifikanzfunktion ( $p \leq .05$ , zweiseitig, Abbildung 29). Auch in dieser Analyse-Variante kann ein Bias im nichtsignifikanten Bereich entlang der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) nicht ausgeschlossen werden. Dabei kann vermutet werden, dass das Bias auf ein Bemühen hindeutet, nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) durch eine Vergrößerung der Stichprobe in den signifikanten Bereich zu „drücken“ (vgl. Furguson, 2009; Abbildung 28).

### *Moderatoranalysen:*

Zum Zusammenhang zwischen Gf und O beträgt die Varianz des artefaktkorrigierten Effektes ausnahmslos weniger als 75%. Dementsprechend wurden Moderatoranalysen durchgeführt. Die kognitiven Leistungstests berücksichtigend, wurden die Studien, in denen der BRT zur Erfassung der fluiden Intelligenzausprägung eingesetzt wurde, mit den Studien verglichen, in denen andere kogLT als der BRT verwendet wurden. Darüber hinaus wurden die Studien, in denen Raven-Tests verwendet wurden, mit den Studien, in denen andere kogLt als die Raven-Tests eingesetzt wurden, verglichen (Tabelle 44). Darüber hinaus wurden Moderatoranalysen hinsichtlich der eingesetzten Persönlichkeitsinventare durchgeführt. Dabei wurden folgende Persönlichkeitsinventare verglichen:

- NEO-Inventare (NEO-PI-(R) oder NEO-FFI) vs.
- NEO-PI-(R) vs.
- NEO-FFI.

In Tabelle 44 sind die metaanalytischen Kennwerte der entsprechenden Moderatoranalysen zum Zusammenhang zwischen Gf und O aufgeführt.

**Tabelle 44**  
Moderatoranalysen zum Zusammenhang zwischen Gf und O

| Gf/O                    | Moderator  | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | % Var. | 90% CV | $\bar{r}$ | $SD_r$ | % Var |
|-------------------------|------------|----------|----------|--------|-------------|--------|--------|-----------|--------|-------|
| <u><i>kogLT</i></u>     |            |          |          |        |             |        |        |           |        |       |
| Autor(en) <sup>1</sup>  | BRT        | 1,681    | 12       | .11    | .059        | 72.39  | .03    | .09       | .058   | 68.14 |
| Autor(en) <sup>1</sup>  | ohne BRT   | 26,354   | 16       | .06    | .050        | 31.18  | -.01   | .04       | .041   | 26.39 |
| Autor(en) <sup>1a</sup> | ohne BRT   | 6,179    | 15       | .09    | .063        | 36.92  | .01    | .09       | .057   | 42.50 |
| CHC <sup>2</sup>        | BRT        | 1,681    | 12       | .11    | .059        | 72.39  | .03    | .09       | .058   | 68.14 |
| CHC <sup>2</sup>        | ohne BRT   | 37,818   | 39       | .08    | .062        | 31.43  | .00    | .06       | .052   | 27.36 |
| CHC <sup>2a</sup>       | ohne BRT   | 17,643   | 38       | .10    | .064        | 39.25  | .02    | .10       | .052   | 44.84 |
| CHC <sup>2</sup>        | Raven      | 23,254   | 13       | .05    | .071        | 20.75  | -.04   | .03       | .047   | 20.05 |
| CHC <sup>2a</sup>       | Raven      | 3,079    | 12       | .12    | .124        | 27.85  | -.01   | .13       | .088   | 33.05 |
| CHC <sup>2</sup>        | ohne Raven | 16,245   | 38       | .10    | .047        | 55.44  | .04    | .09       | .040   | 58.81 |
| <u><i>PSK</i></u>       |            |          |          |        |             |        |        |           |        |       |
| Autor(en) <sup>1</sup>  | NEO        | 27,041   | 25       | .06    | .053        | 37.67  | -.01   | .04       | .044   | 31.93 |
| Autor(en) <sup>1a</sup> | NEO        | 6,866    | 24       | .09    | .067        | 43.76  | .00    | .09       | .063   | 46.23 |
| Autor(en) <sup>1</sup>  | NEO-PI-(R) | 25,256   | 17       | .05    | .039        | 44.63  | .00    | .03       | .036   | 34.74 |
| Autor(en) <sup>1a</sup> | NEO-PI-(R) | 5,081    | 16       | .08    | .050        | 50.77  | .01    | .08       | .058   | 47.83 |
| Autor(en) <sup>1</sup>  | NEO-FFI    | 1,785    | 8        | .20    | .084        | 60.02  | .09    | .12       | .068   | 48.48 |
| CHC <sup>2</sup>        | NEO        | 37,100   | 44       | .08    | .059        | 36.67  | .00    | .05       | .051   | 31.10 |
| CHC <sup>2a</sup>       | NEO        | 10,925   | 43       | .10    | .083        | 44.34  | .02    | .09       | .053   | 47.14 |
| CHC <sup>2</sup>        | NEO-PI-(R) | 32,940   | 29       | .07    | .046        | 41.81  | .01    | .05       | .043   | 31.94 |
| CHC <sup>2a</sup>       | NEO-PI-(R) | 12,765   | 28       | .09    | .046        | 52.87  | .03    | .09       | .046   | 50.79 |
| CHC <sup>2</sup>        | NEO-FFI    | 4,160    | 15       | .15    | .100        | 39.35  | .02    | .11       | .067   | 44.29 |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und –überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; % Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, in denen eine Zuordnung von Gf explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von Gf explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von De Fruyt et al. (2006, *n* = 20,175) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde und die Studie von De Fruyt et al. (2006, *n* = 20,175) exkludiert ist; kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; BRT = Baddeley Reasoning Test (hierbei die Studien inkludiert, die den BRT zur Erfassung der fluiden Intelligenzausprägung verwendet haben; ohne BRT = die Studien inkludiert, die zur Erfassung der fluiden Intelligenz andere kogLT als den BRT verwendet haben; Raven = die Studien inkludiert, die einen Raven Test zur Erfassung der fluiden Intelligenzausprägung verwendet haben; ohne Raven = die Studien inkludiert, die zur Erfassung der fluiden Intelligenz andere kogLT als den Raven Test verwendet haben; NEO = die Studien inkludiert, in denen ausschließlich Costa und McCrae-Inventare (NEO-PI, NEO-PI-R, NEO-FFI) eingesetzt wurden; NEO-PI-(R) = die Studien inkludiert, in denen der NEO-PI-R und der NEO-PI verwendet wurde; NEO-FFI = die Studien berücksichtigt, in denen der NEO-FFI eingesetzt wurde.

Mit keinem gerechneten Moderator konnten mehr als 72.39% der Varianz erklärt werden (Tabelle 44). Es kann davon ausgegangen werden, dass weitere, in der vorliegenden Arbeit nicht identifizierte Moderatoren existieren, die den Zusammenhang zwischen Gf und O erklären.

*Zusammenfassung zum Zusammenhang zwischen der Fluiden Intelligenz und Offenheit für Erfahrung:*

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien vorgenommen wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = .06$ . Dabei kann ein Nulleffekt in der Population nicht ausgeschlossen werden, da das Glaubwürdigkeitsintervall die Null mit einschließt. Nach der Exklusion der Studie von De Fruyt et al. (2006;  $n = 20,175$ ) beträgt die korrigierte Effektgröße  $\rho = .09$ . Hierbei kann von einem positiven Effekt in der Population ausgegangen werden, da die Null außerhalb des Glaubwürdigkeitsintervalls liegt. Unabhängig von der Exklusion bzw. Inklusion der Studie von De Fruyt et al. (2006) kann nicht ausgeschlossen werden, dass weitere Moderatoren existieren, die den Zusammenhang zwischen Gf und O erklären, da die erklärte Varianz kleiner als 75% ist.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, beträgt die korrigierte Effektgröße  $\rho = .08$ . Dabei kann ein Nulleffekt in der Population nicht ausgeschlossen werden, da das Glaubwürdigkeitsintervall auch in dieser Analyse-Variante die Null mit einschließt. Nach der Exklusion der Studie von De Fruyt et al. (2006,  $n = 20,175$ ) beträgt die korrigierte Effektgröße  $.10$  und es kann von einem positiven Effekt ausgegangen werden, da das Glaubwürdigkeitsintervall die Null nicht einschließt. Die erklärte Varianz beträgt in dieser Variante ebenfalls deutlich weniger als 75%.

Hinsichtlich eines Publikationsbias kann nach der „trim- and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien durchgeführt wurde, aufgrund der Höhe des Schätzers mit  $R_0 = 1$  und  $L_0 = 3.55$  von einer asymmetrischen Verteilung der publizierten Effektgrößen ausgegangen werden. Demnach ist ein Publikationsbias nicht auszuschließen. In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, kann nach der „trim- and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) auf Grund der Höhe des Schätzers mit  $R_0 < 1$  von einer symmetrischen Verteilung der publizierten Effektgrößen ausgegangen und ein Publikationsbias ausgeschlossen werden. Der zweite Schätzers zeigte mit  $L_0 = 1.80$  eine asymmetrische Verteilung der publizierten

Effektgrößen an. Ein Publikationsbias war in dem Fall nicht auszuschließen. Nach den durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens, kann in beiden Analyse-Varianten (Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien vs. CHC-Theorie) ein Publikationsbias nicht ausgeschlossen werden. Dabei kann davon ausgegangen werden, dass vermehrt signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Vergleich zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt sind. Hinsichtlich der Moderatoranalysen, in denen ausnahmslos weniger als 75% der Varianz erklärt werden konnten, ist davon auszugehen, dass weitere Moderatoren zum Zusammenhang zwischen Gf und O existieren, die allerdings im Rahmen der vorliegenden Arbeit nicht identifiziert wurden.

### 5.9 Ergebnisse zur Kristallisierten Intelligenz und Offenheit für Erfahrung

*Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien:*

In dieser Analyse-Variante beträgt die inkludierte Studienanzahl  $k = 13$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .14$  ( $SD = .03$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 5,912$  (Range:  $n = 64$  bis  $71,887$ ;  $SD = 19,828.48$ ;  $N = 76,867$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. -überschätzung korrigierte Effektgröße beträgt  $\rho = .20$  ( $SD_{\rho} = .04$ ). Das Glaubwürdigkeitsintervall schließt die Null nicht mit ein und die erklärte Varianz beträgt 15.92%. (Tabelle 45).

Nach der anschließenden Exklusion der Studie von Reeve et al. (2006) mit einer Stichprobengröße von  $n = 71,887$  beträgt die unkorrigierte Effektgröße  $\bar{r}_{bare-bones} = .18$  ( $SD = .11$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 415$  (Range:  $n = 64$  bis  $1,821$ ;  $SD = 497.08$ ;  $N = 4,980$ ). Die artefaktkorrigierte Effektgröße beträgt  $\rho = .24$  ( $SD_{\rho} = .15$ ). Die Null liegt hierbei ebenfalls nicht im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 15.37% (Tabelle 45).

Unabhängig von Exklusion bzw. Inklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) kann von einem positiven Effekt in der Population ausgegangen werden, da das Glaubwürdigkeitsintervall die Null nicht mit einschließt. Allerdings sollte die Höhe des Populationszusammenhangs nicht generalisiert werden, da die erklärte Varianz deutlich kleiner als 75% ist. Dementsprechend wurden Moderatoranalysen durchgeführt (siehe unten).

**Tabelle 45**

Metaanalytische Ergebnisse zum Zusammenhang zwischen Gc und O

| Gc/Offenheit                         | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | %Var. | 90% CV | $\bar{r}$ | $SD_r$ | %Var  |
|--------------------------------------|----------|----------|--------|-------------|-------|--------|-----------|--------|-------|
| kodiert nach Autor(en) <sup>1</sup>  | 76,867   | 13       | .20    | .040        | 15.92 | .15    | .14       | .028   | 17.14 |
| kodiert nach Autor(en) <sup>1a</sup> | 4,980    | 12       | .24    | .149        | 15.37 | .05    | .18       | .105   | 17.00 |
| kodiert nach CHC <sup>2</sup>        | 81,369   | 34       | .21    | .057        | 18.59 | .13    | .15       | .043   | 17.71 |
| kodiert nach CHC <sup>2a</sup>       | 9,482    | 33       | .29    | .127        | 23.43 | .13    | .23       | .093   | 26.78 |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und -überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; %Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße; <sup>1</sup> = Datensätze, in denen eine Zuordnung von Gc explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von Gc explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist.

*Zuordnung der Fähigkeitskonstrukte gemäß CHC-Theorie:*

In dieser Analyse-Variante beträgt die inkludierte Studienzahl *k* = 34 bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .15$  ( $SD = .04$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 2,393$  (Range: *n* = 50 bis 71,887;  $SD = 12,280.93$ ; *N* = 81,369). Die artefaktkorrigierte Effektgröße beträgt  $\rho = .21$  ( $SD_{\rho} = .06$ ). Die Null liegt nicht im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 18.59%.

Nach der anschließenden Exklusion der Studie von Reeve et al. (2006, *n* = 71,887) erhöht sich die die unkorrigierte Effektgröße von  $\bar{r}_{bare-bones} = .15$  ( $SD = .04$ ) auf  $\bar{r}_{bare-bones} = .23$  ( $SD = .09$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 287$  (*n* = Range: 50 bis 789;  $SD = 206.65$ ; *N* = 9,482). Die artefaktkorrigierten Effektgröße erhöht sich durch die Exklusion der Studie von Reeve et al. (2006) von  $\rho = .21$  ( $SD_{\rho} = .06$ ) auf  $\rho = .29$  ( $SD_{\rho} = .13$ ). Die Null liegt ebenfalls nicht im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 23.43% (Tabelle 45).

Unabhängig von der Inklusion bzw. Exklusion der Studie von Reeve et al. (2006, *n* = 71,887) kann auch in dieser Analyse-Variante von einem positiven Effekt in der Population ausgegangen werden, da die Null nicht im Glaubwürdigkeitsintervall liegt. Auch in dieser Analyse-Variante wurden Moderatoranalysen, aufgrund der erklärten Varianz, die deutlich kleiner als 75% ist, durchgeführt.



*Untersuchung des Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe des Schätzer mit  $R_0 = 1$  von einer asymmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .24$  ausgegangen und ein Publikationsbias nicht ausgeschlossen werden. Nach der Exklusion einer Effektgröße mit  $r = .02$  kann eine symmetrische Verteilung der mittlere, neu bestimmten Effektgröße von  $\bar{r} = .25$  angenommen und ein Publikationsbias ausgeschlossen werden. Der zweite Schätzer zeigt hingegen mit  $L_0 < 1$  eine Symmetrie hinsichtlich der Summer der Ränge publizierter Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .24$  an. Ein Publikationsbias wäre demnach auszuschließen. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der sechs relevanten Stichprobengrößen real publizierter Studien und die entsprechende Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 46 aufgeführt.

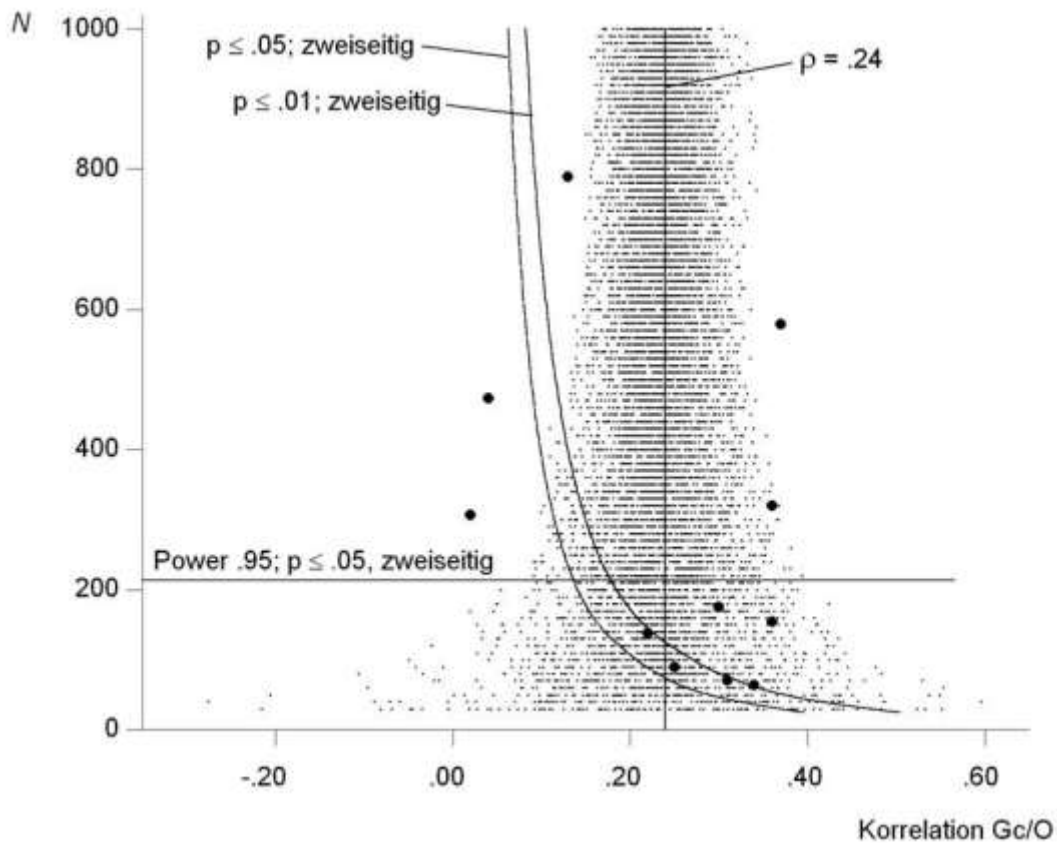
**Tabelle 46**

Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | $N$       | $k$ (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | $k$ (Simulation) <sup>3</sup> |
|-----------|-----------|------------------------------|-----------------------|-------------------------------|
| 1.        | 60 – 100  | 3                            | 50.00                 | 2,500                         |
| 2.        | 110 – 150 | 2                            | 33.33                 | 1,667                         |
| 3.        | 160 – 200 | 1                            | 16.67                 | 834                           |
| $\Sigma$  |           | <b>6</b>                     | <b>100.00</b>         | <b>5,001<sup>4</sup></b>      |

*Anmerkungen.*  $p = .24$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten  $k$  für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt  $k$  von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt  $k = 5,001$ .

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .69 hat sich aus einem Verhältnis von 3,467 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 1,534 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Da die simulierten Effektgrößen keinerlei Publikationsbias aufweisen, kann hierbei ein Verhältnis von 69% signifikanten ( $p \leq .05$ , zweiseitig) zu 31% nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) als Referenz zur Beurteilung eines Publikationsbias herangezogen werden. Der aus den tatsächlich publizierten Effektgrößen berechnete Index beträgt eins (sechs signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; keine nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 30).



**Abbildung 30.**  $\rho = .24$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 46) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig) und  $p \leq .01$  (zweiseitig); GPower (Faul & Erdfelder, 1992; Power .95,  $p \leq .05$ , zweiseitig):  $n = 215$ ; Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien.

Um bei den tatsächlich publizierten Effektgrößen einen Index von .69 – ohne Publikationsbias – zu erreichen, müssten drei nichtsignifikante Effektgrößen ( $p \leq .05$ , zweiseitig) in einem Verhältnis von sechs signifikanten ( $p \leq .05$ , zweiseitig) zu drei nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) unter den gegebenen Parametern publiziert sein. Dementsprechend kann tendenziell von einem Publikationsbias ausgegangen werden: Zu viele signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) sind im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt. Allerdings sollte in dem Fall die geringe Studienanzahl von  $k = 6$  berücksichtigt werden und die Annahme eines Publikationsbias unter Vorbehalt vorgenommen werden.

Anhand der Abbildung 30 ist auffällig, dass sich vier signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) entlang der eingezeichneten Signifikanzfunktionen mit  $p \leq .05$  (zweiseitig) und  $p \leq .01$  (zweiseitig) verteilen. Demgegenüber sind keine publizierten Effektgrößen mit  $p > .05$  (zweiseitig)

entlang der eingezeichneten Signifikanzfunktionen publiziert. Offensichtlich tritt auch hier der Effekt eines Bias im nichtsignifikanten Bereich entlang der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) auf. Dabei kann vermutet werden, dass das Bias auf ein Bemühen hindeutet, nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) durch eine Vergrößerung der Stichprobe in den signifikanten Bereich zu „drücken“.

Darüber hinaus fallen zwei publizierte Effektgrößen auf, die im linken Bereich, deutlich außerhalb der simulierten Effektgrößen publiziert sind (Abbildung 30). In Tabelle 47 sind zu den zwei Effektgrößen studienspezifische Merkmale aufgeführt. Die zwei Studien von Arteché et al. (2008) sowie Chamorro-Premuzic und Arteché (2008) kommen aus der selben Forschergruppe um Adrian Furnham aus Großbritannien, wurden im gleichen Jahr publiziert und haben zur Erfassung von O den NEO-PI-R eingesetzt. Allerdings können aufgrund der Gemeinsamkeiten keine Auffälligkeiten gefunden werden, weshalb die publizierten Effektgrößen deutlich von den simulierten Effektgrößen abweichen.

**Tabelle 47**

Deskriptive Angaben publizierter Studien, in denen die Effektgröße des Zusammenhangs zwischen  $G_c$  und 0 von den gezogenen Effektgrößen aus der adaptierten Simulation abweicht

| Studie   | kogLT | PSK      | Land | Alter                          | Stichprobe                   | Thema  |
|--|-------|----------|------|--------------------------------|------------------------------|--|
| Arteché, Chamorro-Premuzic, Furnham und Crump (2008) | WGCTA | NEO-PI-R | GB   | $M = 43.91$<br>( $SD = 7.90$ ) | Angestellte<br>( $n = 307$ ) | Emotionale Intelligenz und Persönlichkeit      |
| Chamorro-Premuzic und Arteché (2008)                 | MHT   | NEO-PI-R | GB   | $M = 20.31$<br>( $SD = 3.43$ ) | Studenten<br>( $n = 473$ )   | Längsschnittstudie zur akademischen Performanz |

Anmerkungen. kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; WGCTA = Watson Glaser Critical Thinking Appraisal; MHT = Mill-Hill-Test.

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durchgeführt wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe der Schätzer mit  $R_0 < 1$  eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .26$  angenommen und ein Publikationsbias ausgeschlossen werden. Hingegen kann aufgrund der Höhe des zweiten Schätzers mit  $L_0 = 1.2$  von einer Asymmetrie hinsichtlich der Summe der Ränge publizierter Effektgrößen ausgegangen und ein Publikationsbias nicht ausgeschlossen werden. Nach der anschließenden Exklusion von zwei Effektgrößen ( $L_0^+ = 2$ ) mit positiven Vorzeichen und den größten Beträgen kann eine symmetrische Verteilung publizierter Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .25$  mit

$L_0 < 1$  angenommen und ein Publikationsbias ausgeschlossen werden. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 13 relevanten Stichprobengrößen real publizierter Studien und die entsprechende Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 48 aufgeführt.

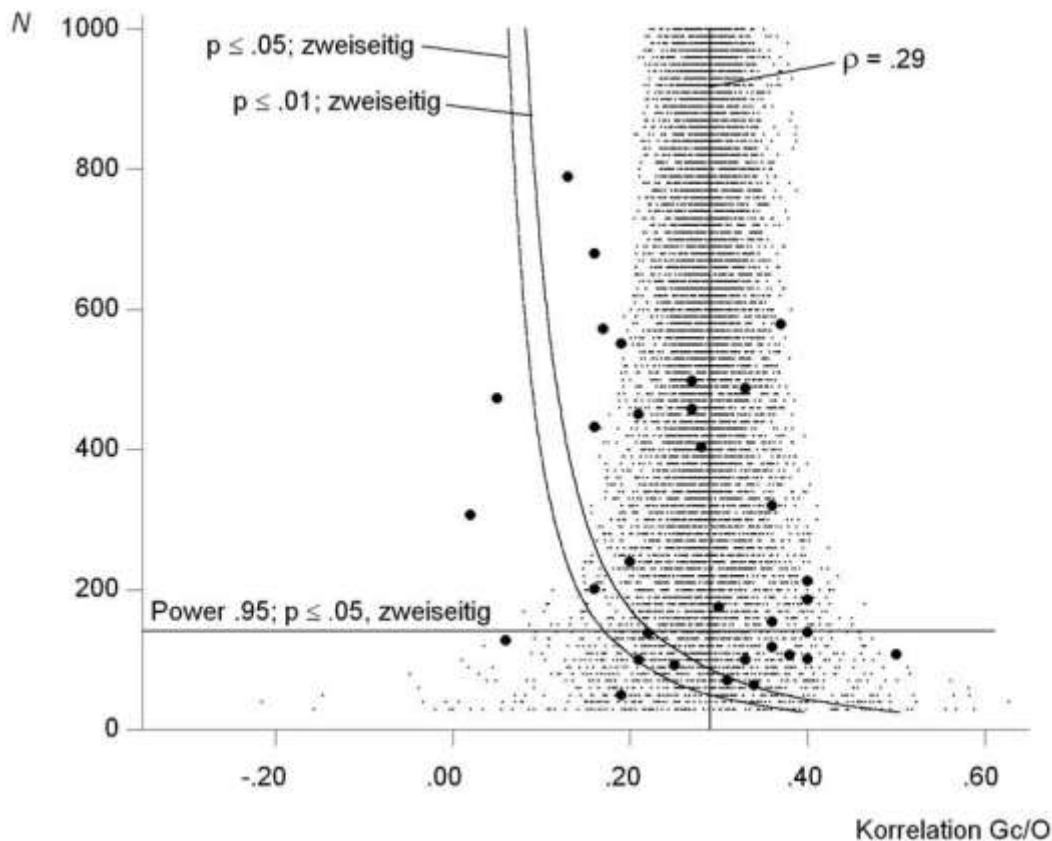
**Tabelle 48**

Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | $N$       | $k$ (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | $k$ (Simulation) <sup>3</sup> |
|-----------|-----------|------------------------------|-----------------------|-------------------------------|
| 1.        | 10 – 50   | 1                            | 7.69                  | 385                           |
| 2.        | 60 – 100  | 6                            | 46.15                 | 2,310                         |
| 3.        | 110 – 150 | 6                            | 46.15                 | 2,310                         |
| $\Sigma$  |           | <b>13</b>                    | <b>100.00</b>         | <b>5,005<sup>4</sup></b>      |

Anmerkungen.  $p = .29$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten  $k$  für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt  $k$  von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt  $k = 5,005$ .

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .80 hat sich aus einem Verhältnis von 3,984 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 1,021 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Da die simulierten Effektgrößen keinerlei Publikationsbias aufweisen, kann hierbei ein Verhältnis von 80% signifikanten ( $p \leq .05$ , zweiseitig) zu 20% nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) als Referenz zur Beurteilung eines Publikationsbias herangezogen werden. Der aus den tatsächlich publizierten Effektgrößen berechnete Index beträgt .85 (11 signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; zwei nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 31).



**Abbildung 31.**  $\rho = .29$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 48) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig) und  $p \leq .01$  (zweiseitig); GPower (Faul & Erdfelder, 1992; Power .95,  $p \leq .05$ , zweiseitig):  $n = 144$ ; Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie (McGrew, 2009).

Um bei den tatsächlich publizierten Effektgrößen einen Index von ungefähr .80 – ohne Publikationsbias – zu erreichen, müsste eine weitere nichtsignifikante Effektgrößen ( $p \leq .05$ , zweiseitig) in einem Verhältnis von 11 signifikanten ( $p \leq .05$ , zweiseitig) zu drei nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) unter den gegebenen Parametern publiziert sein. Da allerdings nicht ausgeschlossen werden kann, dass eine weitere nichtsignifikante Effektgröße ( $p > .05$ , zweiseitig) zum Zusammenhang zwischen Gc und O mit  $n \leq 144$  existiert, kann ein Publikationsbias ausgeschlossen werden.

Kann insgesamt ein Publikationsbias ausgeschlossen werden, fällt auf, dass mindestens sechs signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) entlang der eingezeichneten Signifikanzfunktionen ( $p \leq .05$ , zweiseitig;  $p \leq .01$ , zweiseitig) publiziert sind. Demgegenüber sind keine Effektgrößen mit  $p > .05$  (zweiseitig) entlang der eingezeichneten Signifikanzfunktion mit  $p > .05$  (zweiseitig) publiziert (Abbildung 31). Darüber hinaus sind vier Effektgrößen deutlich außerhalb der

simulierten Effektgrößen publiziert. In Tabelle 49 sind zu den vier Effektgrößen, studienspezifische Merkmale aufgeführt. Allerdings ließen sich keine eindeutigen Auffälligkeiten bzw. Gemeinsamkeiten feststellen, weshalb die publizierten Effektgrößen von den simulierten Effektgrößen im rechten äußeren Bereich abweichen.

**Tabelle 49**

Deskriptive Angaben publizierter Studien, in denen die Effektgröße des Zusammenhangs zwischen Gc und O von den gezogenen Effektgrößen aus der adaptierten Simulation abweicht

| Studie   | kogLT                | PSK      | Land | Alter                           | Stichprobe                          | Thema   |
|--|----------------------|----------|------|---------------------------------|-------------------------------------|---|
| Arteche, Chamorro-Premuzic, Furnham und Crump (2008) | WGCTA                | NEO-PI-R | GB   | $M = 43.91$<br>( $SD = 7.90$ )  | Angestellte<br>( $n = 307$ )        | Emotionale Intelligenz und Persönlichkeit       |
| Chamorro-Premuzic und Arteche (2008)                 | MHT                  | NEO-PI-R | GB   | $M = 20.31$<br>( $SD = 3.43$ )  | Studenten<br>( $n = 473$ )          | Längstschnittstudie zur akademischen Performanz |
| Beauducel, Liepmann, Felte und Nettelstroth (2007)   | IST-2000R            | NEO-FFI  | D    | $M = 18.00$<br>( $SD = 1.38$ )  | Schüler<br>( $n = 789$ )            | akademische Fähigkeiten                         |
| Zimprich, Allemann und Dellenbach (2009)             | Teile aus dem WAIS-R | NEO-FFI  | D    | $M = 43.72$<br>( $SD = 0.93.$ ) | nicht spezifiziert<br>( $n = 679$ ) | Beziehung zwischen Gf/Gc und O                  |

*Anmerkungen.* kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; WGCTA = Watson Glaser Critical Thinking Appraisal; MHT = Mill-Hill-Test; IST-2000R = Intelligenz-Struktur-Test 2000R; WAIS-R = Wechsler-Intelligenz Test, revidierte Fassung.

*Moderatoranalysen:*

Zum Zusammenhang zwischen Gc und O beträgt die Varianz des artefaktkorrigierten Effektes ausnahmslos weniger als 75%. Dementsprechend wurden Moderatoranalysen durchgeführt. Die kognitiven Leistungstests berücksichtigend, wurden die Studien, in denen der GKT zur Messung von Gc verwendet wurde, mit den Studien verglichen, in denen andere kogLT als der GKT verwendet wurden (bspw. Watson-Glaser Critical Thinking Appraisal).

Darüber hinaus wurden Moderatoranalysen hinsichtlich der eingesetzten Persönlichkeitsinventare durchgeführt. Dabei wurden folgende Persönlichkeitsinventare verglichen:

- FFM-Inventare (bspw. NEO-PI-(R), NEO-FFI, Personal Characteristics Inventory)
- NEO-PI-(R) vs.
- NEO-FFI vs.
- FFM-Inventare, allerdings keine NEO-Inventare.

In Tabelle 50 sind die metaanalytischen Kennwerte der entsprechenden Moderatoranalysen, bei denen die inkludierte Studienanzahl mindestens fünf beträgt (vgl. Kapitel 4.5), aufgeführt.

**Tabelle 50**

Moderatoranalysen zum Zusammenhang zwischen Gc und O

| Gc/O                   | Moderator   | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | % Var. | 90% CV | $\bar{r}$ | $SD_r$ | % Var |
|------------------------|-------------|----------|----------|--------|-------------|--------|--------|-----------|--------|-------|
| <u>kogLT</u>           |             |          |          |        |             |        |        |           |        |       |
| CHC <sup>1</sup>       | GKT         | 832      | 6        | .32    | .149        | 22.37  | .13    | .33       | .087   | 42.89 |
| CHC <sup>1</sup>       | ohne GKT    | 80,537   | 28       | .20    | .052        | 18.60  | .14    | .15       | .038   | 18.50 |
| CHC <sup>1a</sup>      | ohne GKT    | 8,650    | 27       | .29    | .124        | 23.40  | .13    | .22       | .088   | 27.00 |
| <u>PSK</u>             |             |          |          |        |             |        |        |           |        |       |
| Autor(en) <sup>2</sup> | NEO         | 4,404    | 11       | .20    | .122        | 22.14  | .05    | .15       | .085   | 24.93 |
| Autor(en) <sup>2</sup> | NEO-PI-(R)  | 3,138    | 8        | .16    | .072        | 42.77  | .07    | .13       | .061   | 40.13 |
| CHC <sup>1</sup>       | NEO         | 7,798    | 28       | .27    | .125        | 24.32  | .11    | .22       | .095   | 26.68 |
| CHC <sup>1</sup>       | NEO-PI-(R)  | 3,445    | 15       | .23    | .128        | 22.93  | .07    | .21       | .109   | 25.24 |
| CHC <sup>1</sup>       | NEO-FFI     | 4,353    | 13       | .32    | .098        | 34.62  | .19    | .23       | .082   | 28.86 |
| CHC <sup>1</sup>       | FFM, notNEO | 2,363    | 6        | .36    | .089        | 35.30  | .25    | .25       | .069   | 31.63 |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und -überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; % Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde; <sup>1a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc gemäß der Autoren publizierter Studien berücksichtigt wurde; kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; GKT = General Knowledge Test; die Studien inkludiert, die den GKT zur Erfassung der kristallisierten Intelligenzausprägung verwendet haben; ohne GKT = die Studien inkludiert, die zur Erfassung der kristallisierten Intelligenz andere kogLT als den GKT verwendet haben; NEO = die Studien inkludiert, in denen der NEO-PI-(R) und der NEO-FFI berücksichtigt wurden; NEO-PI-(R) = die Studien inkludiert, in denen der NEO-PI-R und der NEO-PI verwendet wurde; NEO-FFI = die Studien berücksichtigt, in denen der NEO-FFI eingesetzt wurde; FFM, no NEO = die Studien inkludiert, in denen „Big-Five“-Inventare eingesetzt wurden, allerdings nicht von Costa und McCrae.

Mit keinem Moderator konnten mehr als 42.77% der Varianz erklärt werden (Tabelle 44). Es kann davon ausgegangen werden, dass weitere, in der vorliegenden Arbeit nicht identifizierte Moderatoren existieren, die den Zusammenhang zwischen Gc und O erklären.

*Zusammenfassung zum Zusammenhang zwischen der Kristallisierten Intelligenz und Offenheit für Erfahrung:*

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = .20$ . Nach der anschließenden Exklusion der Studie von Reeve et al. (2006, *n* =71,887) beträgt die korrigierte Effektgröße  $\rho = .24$ . Unabhängig von der Exklusion bzw. Inklusion der Studie von

Reeve et al. (2006) kann von einem positiven Effekt in der Population ausgegangen werden, da das Glaubwürdigkeitsintervall die Null nicht mit einschließt. Allerdings sollte die Höhe der artefaktkorrigierten Effektgröße nicht generalisiert werden, da die erklärte Varianz deutlich kleiner als 75% ist.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, beträgt die korrigierte Effektgröße  $\rho = .21$ . Nach der Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) beträgt  $\rho = .29$ . Unabhängig von der Exklusion bzw. Inklusion der Studie von Reeve et al. (2006) kann auch in dieser Analyse-Variante von einem positiven Effekt in der Population ausgegangen werden. Moderatoranalysen wurden aufgrund der erklärten Varianz, die deutlich kleiner als 75% war, ebenfalls durchgeführt.

Hinsichtlich eines Publikationsbias, kann nach der „trim- and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien durchgeführt wurde, ein Publikationsbias mit  $R_0 = 1$  nicht ausgeschlossen werden. Hingegen kann mit  $L_0 < 1$  ein Publikationsbias ausgeschlossen werden. Nach der durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens, kann tendenziell von einem Publikationsbias ausgegangen werden. Allerdings sollte eine Einschätzung zum Publikationsbias aufgrund der geringen Studienanzahl von  $k = 6$  unter Vorbehalt vorgenommen werden.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, kann aufgrund der Höhe des Schätzers mit  $R_0 < 1$  ein Publikationsbias ausgeschlossen werden. Mit  $L_0 = 1.2$  kann hingegen ein Publikationsbias nicht ausgeschlossen werden. Nach der durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens, kann ein Publikationsbias ausgeschlossen werden.

Mit keinem Moderator konnten mehr als 42.77% der Varianz erklärt werden. Es ist davon auszugehen, dass weitere, im Rahmen der vorliegenden Arbeit nicht identifizierte Moderatoren existieren, die den Zusammenhang zwischen Gf und O erklären.



### 5.10 Ergebnisse zur Allgemeinen Intelligenz und Verträglichkeit

*Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien:*

In dieser Variante beträgt die inkludierte Studienanzahl  $k = 57$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .08$  ( $SD = .05$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 1,518$  (Range:  $n = 59$  bis  $71,887$ ;  $SD = 9,490.88$ ;  $N = 86,499$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. –überschätzung korrigierte Effektgröße beträgt  $\rho = .10$  ( $SD_{\rho} = .07$ ). Das Glaubwürdigkeitsintervall schließt die Null mit ein und die erklärte Varianz beträgt 18.15%. (Tabelle 51).

Nach der anschließenden Exklusion der Studie von Reeve et al. (2006) mit einer Stichprobengröße von  $n = 71,887$  beträgt die unkorrigierte Effektgröße  $\bar{r}_{bare-bones} = .00$  ( $SD = .10$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 261$  (Range:  $n = 59$  bis  $1,886$ ;  $SD = 270.18$ ;  $N = 14,612$ ). Die artefaktkorrigierte Effektgröße verringert sich durch die Exklusion der Studie von Reeve et al. (2006) von  $\rho = .10$  ( $SD_{\rho} = .07$ ) auf  $\rho = .01$  ( $SD_{\rho} = .11$ ). Die Null liegt hierbei ebenfalls im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 28.36% (Tabelle 51).

Unabhängig von der Exklusion bzw. Inklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) kann ein Nulleffekt in der Population nicht ausgeschlossen werden, da das Glaubwürdigkeitsintervall die Null mit einschließt. Aufgrund der erklärten Varianz von maximal 28.36% ist davon auszugehen, dass weitere Moderatoren existieren, die den Zusammenhang zwischen g und V erklären. Dementsprechend wurden Moderatoranalysen durchgeführt (siehe unten).

**Tabelle 51**

Metaanalytische Ergebnisse zum Zusammenhang zwischen g und V

| g/Verträglichkeit                    | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | %Var. | 90% CV | $\bar{r}$ | $SD_r$ | %Var  |
|--------------------------------------|----------|----------|--------|-------------|-------|--------|-----------|--------|-------|
| kodiert nach Autor(en) <sup>1</sup>  | 86,499   | 57       | .10    | .073        | 18.15 | .00    | .08       | .052   | 19.34 |
| kodiert nach Autor(en) <sup>1a</sup> | 14,612   | 56       | .01    | .114        | 28.36 | -.15   | .00       | .100   | 27.91 |
| kodiert nach CHC <sup>2</sup>        | 86,847   | 59       | .10    | .072        | 19.82 | .01    | .07       | .052   | 20.54 |
| kodiert nach CHC <sup>2a</sup>       | 14,960   | 58       | .00    | .120        | 31.13 | -.15   | .00       | .093   | 31.41 |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und –überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; %Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße; <sup>1</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von Reeve et al. (2006,  $n = 71,887$ ) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde und die Studie von Reeve et al. (2006,  $n = 71,887$ ) exkludiert ist.

## Kapitel 5 – Ergebnisse

### *Zuordnung der Fähigkeitskonstrukte gemäß CHC-Theorie:*

In dieser Analyse-Variante beträgt die inkludierte Studienzahl  $k = 59$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .07$  ( $SD = .05$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 1,466$  (Range:  $n = 59$  bis  $71,887$ ;  $SD = 9,327.79$ ;  $N = 86,847$ ). Die artefaktkorrigierte Effektgröße beträgt  $\rho = .10$  ( $SD_{\rho} = .07$ ). Die erklärte Varianz beträgt 19.82% und die Null liegt nicht im Glaubwürdigkeitsintervall.

Nach der anschließenden Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) beträgt die unkorrigierte Effektgröße  $\bar{r}_{bare-bones} = .00$  ( $SD = .09$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 252$  (Range:  $n = 59$  bis  $900$ ;  $SD = 181.62$ ;  $N = 14,960$ ). Die artefaktkorrigierte Effektgröße verringert sich hierbei von  $\rho = .10$  ( $SD_{\rho} = .07$ ) auf  $\rho = .00$  ( $SD_{\rho} = .12$ ). Die erklärte Varianz beträgt 31.13% (Tabelle 51).

Auch in dieser Analyse-Variante kann ein Populationseffekt von Null, unabhängig von der Exklusion bzw. Inklusion der Studie von Reeve et al. (2006), nicht ausgeschlossen werden. Darüber hinaus wurden aufgrund der erklärten Varianz, die kleiner als 75% ist, ebenfalls Moderatoranalysen durchgeführt (siehe unten).

### *Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe des Schätzer mit  $R_0 < 1$  von einer symmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .02$  ausgegangen und ein Publikationsbias ausgeschlossen werden. Hinsichtlich des zweiten Schätzers ( $L_0 = 7.06$ ) kann ein Publikationsbias nicht ausgeschlossen werden. Nach der Exklusion von acht Effektgrößen ( $L_0^+ = 8$ ) mit negativen Vorzeichen und den höchsten Beträgen kann von einer symmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .03$  mit  $L_0 < 1$  ausgegangen und ein Publikationsbias ausgeschlossen werden. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 56 Stichprobengrößen real publizierter Studien und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 52 aufgeführt.

**Tabelle 52**

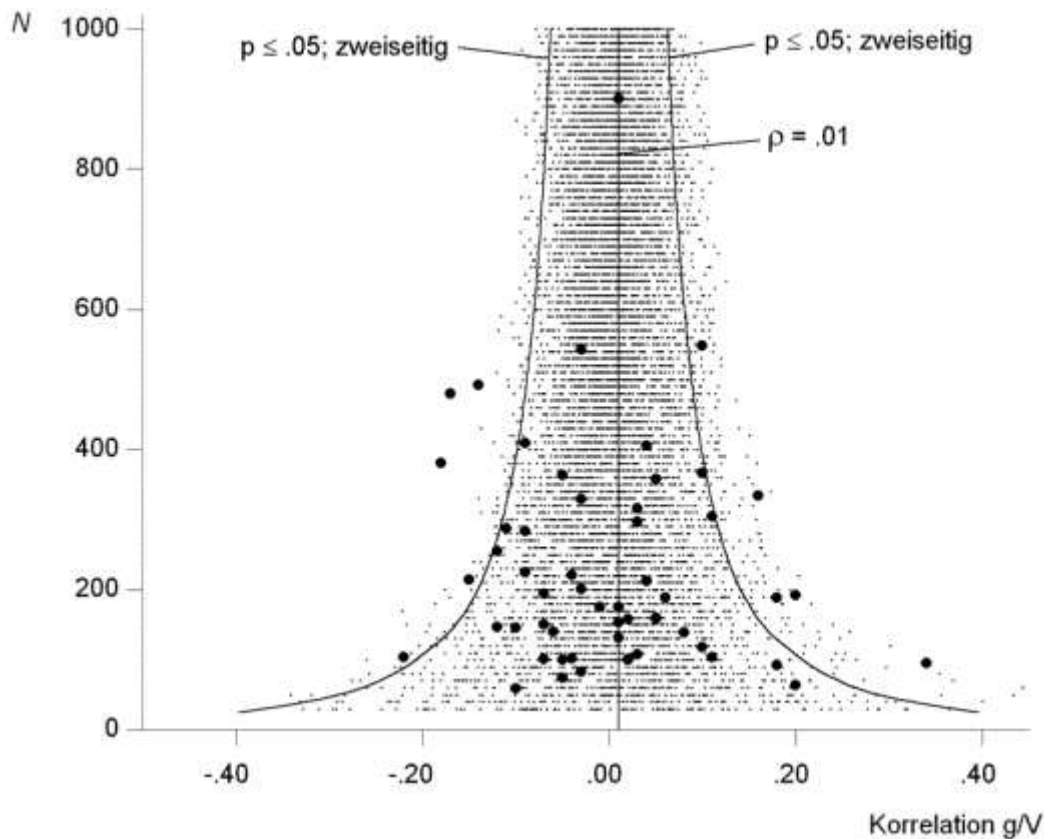
Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | <i>N</i>      | <i>k</i> (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | <i>k</i> (Simulation) <sup>3</sup> |
|-----------|---------------|-----------------------------------|-----------------------|------------------------------------|
| 1.        | 60 – 100      | 13                                | 23.21                 | 1,161                              |
| 2.        | 110 – 150     | 9                                 | 16.07                 | 804                                |
| 3.        | 160 – 200     | 10                                | 17.86                 | 893                                |
| 4.        | 210 – 250     | 4                                 | 7.14                  | 357                                |
| 5.        | 260 – 300     | 4                                 | 7.14                  | 357                                |
| 6.        | 310 – 350     | 4                                 | 7.14                  | 357                                |
| 7.        | 360 – 400     | 4                                 | 7.14                  | 357                                |
| 8.        | 410 – 450     | 2                                 | 3.57                  | 179                                |
| 9.        | 460 – 500     | 2                                 | 3.57                  | 179                                |
| 10.       | 510 – 550     | 2                                 | 3.57                  | 179                                |
| 11.       | 860 – 900     | 1                                 | 1.79                  | 90                                 |
| 12.       | 1,860 – 1,900 | 1                                 | 1.79                  | 90                                 |
| $\Sigma$  |               | <b>56</b>                         | <b>100.00</b>         | <b>5,003<sup>4</sup></b>           |

Anmerkungen.  $p = .01$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten *k* für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt *k* von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt *k* = 5,003.

Unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich hat sich mittels adaptierter Simulation (Kapitel 4.6.7) ein Index von .03 berechnet (165 signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 4,838 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig). Hinsichtlich der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich hat mittels adaptierter Simulation (Kapitel 4.6.7) ein Index von .02 berechnet (79 signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 4,924 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 32).

Der aus den tatsächlich publizierten Effektgrößen berechnete Index unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich beträgt .11 (sechs signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 50 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 32). Der aus den tatsächlich publizierten Effektgrößen berechnete Index unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich beträgt .09 (fünf signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 51 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 32).



**Abbildung 32.**  $\rho = .01$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 52) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktionen im positiven und negativen Wertebereich mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien.

Um bei den tatsächlich publizierten Effektgrößen einen Index von .03 – ohne Publikationsbias – zu erreichen, müssten unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich weitere 120 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von sechs signifikanten ( $p \leq .05$ , zweiseitig) zu 170 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Die Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich berücksichtigend, müssten sogar weitere 152 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von fünf signifikanten ( $p \leq .05$ , zweiseitig) zu 203 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein, um einen Index von .02 – ohne Publikationsbias – zu erreichen.

Entsprechend der Ergebnisse kann anhand der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) sowohl im positiven als auch negativen Wertebereich ein Publikationsbias nicht ausgeschlossen werden. Es sind jeweils zu viele signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu

nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde.

Anhand der Abbildung 32 kann es auffallen, dass im linken Bereich, außerhalb der simulierten Effektgrößen, drei Effektgrößen publiziert sind. In Tabelle 53 sind zu den drei Effektgrößen studienspezifische Merkmale aufgeführt. Allerdings ließen sich keine eindeutigen Auffälligkeiten bzw. Gemeinsamkeiten feststellen, weshalb die publizierten Effektgrößen von den simulierten Effektgrößen abweichen.

**Tabelle 53**

Deskriptive Angaben publizierter Studien, in denen die Effektgröße des Zusammenhangs zwischen  $g$  und  $V$  von den gezogenen Effektgrößen aus der adaptierten Simulation abweicht

| Studie                            | kogLT             | PSK          | Land    | Alter  | Stichprobe                           | Thema   |
|-----------------------------------|-------------------|--------------|---------|--|--------------------------------------|---|
| Taggar, Hackett und Saha (1999)   | WPT               | NEO-PI-R     | CAN     | $M = 21.00$<br>( $SD = n.b.$ )                     | Studenten<br>( $n = 480$ )           | Individuelle Unterschiede im Team im Kontext von Führungsverhaltens |
| Allik und Realo (1997)            | unbekannter kogLT | NEO-PI       | Estland | $M = 19.10$<br>( $SD = 2.30$ )                     | angehende Studenten<br>( $n = 381$ ) | akademische Fähigkeiten   |
| Austin et al. (2002) <sup>4</sup> | Testbatterie,     | short-NEO-PI | SE      | $M = n.b.$<br>( $SD = n.b.$ )<br>Range:<br>55 - 74 | keine Angabe<br>( $n = 405$ )        | Intelligenz und persönliche/soziale Adjustierung                    |

Anmerkungen. kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar.

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durchgeführt wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund des Schätzers mit  $R_0 = 1$  und  $L_0 = 1.41$  von einer asymmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .01$  ausgegangen werden. Ein Publikationsbias kann dementsprechend nicht ausgeschlossen werden. Nach der anschließenden Exklusion einer Effektgröße ( $R_0^+ = 1$ ) bzw. zwei Effektgrößen ( $L_0^+ = 2$ ) kann ein Publikationsbias aufgrund einer symmetrischen Verteilung publizierter Effektgrößen ausgeschlossen werden. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 58 Stichprobengrößen real publizierter Studien und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 54 aufgeführt.

**Tabelle 54**

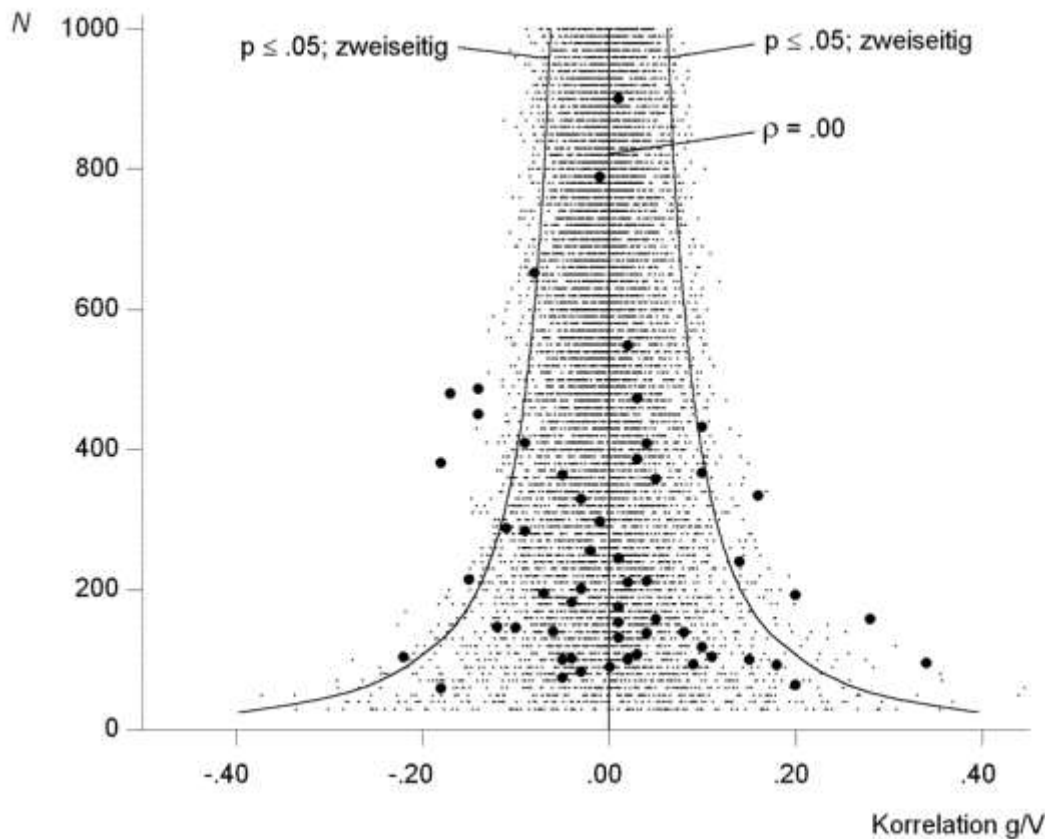
Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | <i>N</i>  | <i>k</i> (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | <i>k</i> (Simulation) <sup>3</sup> |
|-----------|-----------|-----------------------------------|-----------------------|------------------------------------|
| 1.        | 60 – 100  | 15                                | 25.86                 | 1,293                              |
| 2.        | 110 – 150 | 9                                 | 15.52                 | 776                                |
| 3.        | 160 – 200 | 7                                 | 12.07                 | 604                                |
| 4.        | 210 – 250 | 5                                 | 8.62                  | 431                                |
| 5.        | 260 – 300 | 4                                 | 6.90                  | 345                                |
| 6.        | 310 – 350 | 2                                 | 3.45                  | 173                                |
| 7.        | 360 – 400 | 5                                 | 8.62                  | 431                                |
| 8.        | 410 – 450 | 4                                 | 6.90                  | 345                                |
| 9.        | 460 – 500 | 3                                 | 5.17                  | 259                                |
| 10.       | 510 – 550 | 1                                 | 1.72                  | 86                                 |
| 11.       | 610 – 650 | 1                                 | 1.72                  | 86                                 |
| 12.       | 760 – 800 | 1                                 | 1.72                  | 86                                 |
| 13.       | 860 – 900 | 1                                 | 1.72                  | 86                                 |
| <b>Σ</b>  |           | <b>58</b>                         | <b>100.00</b>         | <b>5,001<sup>4</sup></b>           |

Anmerkungen.  $\rho = .00$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten *k* für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt *k* von 5,002; <sup>4</sup> = durch vorgenommene Rundungen beträgt *k* = 5,001.

Der mittels adaptierter Simulation (Kapitel 4.6.7) gebildete Index von .02 hat sich aus einem Verhältnis von 121 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,880 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Da die simulierten Effektgrößen keinerlei Publikationsbias aufweisen, kann ein Verhältnis von 2% signifikanten ( $p \leq .05$ , zweiseitig) zu 97% nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) als Referenz zur Beurteilung eines Publikationsbias herangezogen werden. Da die artefaktkorrigierte Effektgröße  $\rho = .00$  beträgt, gilt diese Referenz für die Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen und positiven Wertebereich.

Die Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich berücksichtigend, beträgt der Index aus den tatsächlich publizierten Effektgrößen .12 (sieben signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 51 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 33). Hingegen beträgt der Index aus den tatsächlich publizierten Effektgrößen unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich .10 (sechs signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 52 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 33).



**Abbildung 33.**  $\rho = .00$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 54) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktionen im positiven und negativen Wertebereich mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie (McGrew, 2009).

Um bei den tatsächlich publizierten Effektgrößen einen Index von .02 – ohne Publikationsbias – zu erreichen, müssten unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich weitere 234 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von sieben signifikanten ( $p \leq .05$ , zweiseitig) zu 285 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich müssten weitere 192 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von sechs signifikanten ( $p \leq .05$ , zweiseitig) zu 244 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein, um einen Index von .02 – ohne Publikationsbias – zu erreichen. Aufgrund der Ergebnisse kann ein Publikationsbias sowohl im positiven als auch negativen Wertebereich nicht ausgeschlossen werden. Es sind zu viele signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde.

## *Kapitel 5 – Ergebnisse*

### *Moderatoranalysen:*

Zum Zusammenhang zwischen  $g$  und  $V$  beträgt die Varianz des artefaktkorrigierten Effektes ausnahmslos weniger als 75%. Dementsprechend wurden Moderatoranalysen durchgeführt. Die kognitiven Leistungstests der publizierten Studien berücksichtigend, wurde die Studien, in denen der WPT zur Erfassung von  $g$  verwendet wurde, mit den Studien verglichen, in denen andere  $kogLT$  als der WPT eingesetzt wurde. Darüber hinaus wurden Moderatoranalysen hinsichtlich der eingesetzten Persönlichkeitsinventare durchgeführt. Dabei wurden folgende Persönlichkeitsinventare verglichen:

- NEO-Inventare vs.
- NEO-PI-(R) vs.
- NEO-FFI vs.
- FFM-Inventare, allerdings keine NEO-Inventare.

In Tabelle 55 sind die metaanalytischen Kennwerte der entsprechenden Moderatoranalysen zum Zusammenhang zwischen  $g$  und  $V$  aufgeführt.



**Tabelle 55**

Moderatoranalysen zum Zusammenhang zwischen g und V

| g/V                     | Moderator   | N      | k  | $\rho$ | $SD_{\rho}$ | % Var. | 90% CV | $\bar{r}$ | $SD_r$ | % Var |
|-------------------------|-------------|--------|----|--------|-------------|--------|--------|-----------|--------|-------|
| <u>kogLT</u>            |             |        |    |        |             |        |        |           |        |       |
| Autor(en) <sup>1</sup>  | WPT         | 4,619  | 26 | -.02   | .171        | 24.00  | .20    | -.01      | .138   | 23.00 |
| Autor(en) <sup>1</sup>  | ohne WPT    | 81,730 | 30 | .11    | .053        | 18.86  | .04    | .08       | .037   | 20.61 |
| Autor(en) <sup>1a</sup> | ohne WPT    | 9,843  | 29 | .02    | .084        | 34.22  | -.09   | .01       | .076   | 33.64 |
| CHC <sup>2</sup>        | WPT         | 4,791  | 26 | -.01   | .111        | 43.22  | .14    | -.01      | .084   | 43.50 |
| CHC <sup>2</sup>        | ohne WPT    | 81,576 | 33 | .11    | .056        | 19.01  | .04    | .08       | .041   | 19.49 |
| CHC <sup>2a</sup>       | ohne WPT    | 9,689  | 32 | .02    | .118        | 28.12  | -.13   | .01       | .092   | 28.28 |
| <u>PSK</u>              |             |        |    |        |             |        |        |           |        |       |
| Autor(en) <sup>1</sup>  | FFM, NEO    | 10,638 | 40 | .01    | .125        | 23.05  | -.15   | .01       | .114   | 22.45 |
| Autor(en) <sup>1</sup>  | NEO-PI-(R)  | 7,789  | 24 | .02    | .090        | 30.48  | -.10   | .01       | .085   | 29.95 |
| Autor(en) <sup>1</sup>  | NEO-FFI     | 2,849  | 16 | .00    | .193        | 18.28  | -.25   | .01       | .159   | 17.27 |
| Autor(en) <sup>1</sup>  | FFM, no NEO | 3,974  | 16 | -.01   | .052        | 72.21  | .06    | -.01      | .039   | 72.94 |
| CHC <sup>2</sup>        | FFM, NEO    | 10,063 | 41 | .00    | .100        | 37.36  | -.13   | .00       | .083   | 37.45 |
| CHC <sup>2</sup>        | NEO-PI-(R)  | 5,740  | 23 | -.01   | .105        | 32.07  | .13    | -.01      | .092   | 32.47 |
| CHC <sup>2</sup>        | NEO-FFI     | 4,323  | 18 | .02    | .088        | 48.52  | -.10   | .02       | .065   | 49.75 |
| CHC <sup>2</sup>        | FFM, no NEO | 4,371  | 16 | -.03   | .096        | 47.79  | .09    | -.03      | .059   | 51.13 |

*Anmerkungen.* N = Stichprobengröße; k = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und –überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; % Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von Reeve et al. (2006, n = 71,887) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde und die Studie von Reeve et al. (2006, n = 71,887) exkludiert ist; kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; WPT = Wonderlic Personal Test (hierbei die Studien inkludiert, die den WPT zur Erfassung der allgemeinen Intelligenzausprägung verwendet haben; ohne WPT = die Studien inkludiert, die zur Erfassung der allgemeinen Intelligenz andere kogLT als den WPT verwendet haben; FFM, NEO = die Studien inkludiert, in denen Costa und McCrae-Inventare (NEO-PI, NEO-PI-R, NEO-FFI) eingesetzt wurden; NEO-PI-(R) = die Studien inkludiert, in denen der NEO-PI-R und der NEO-PI verwendet wurde; NEO-FFI = die Studien berücksichtigt, in denen der NEO-FFI eingesetzt wurde; FFM, no NEO = die Studien berücksichtigt, in denen FFM-Inventare, allerdings keine Costa und McCrae Inventare verwendet wurden.

Mit keinem Moderator konnten mehr als 72.21% der Varianz erklärt werden (Tabelle 55). Es kann nicht ausgeschlossen werden, dass weitere, in der vorliegenden Arbeit nicht identifizierte Moderatoren existieren, die den Zusammenhang zwischen g und V erklären.

*Zusammenfassung zum Zusammenhang zwischen der Allgemeinen Intelligenz und Verträglichkeit:*

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien vorgenommen wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = .10$ . Nach der Exklusion der Studie von Reeve et al. (2006) mit einer Stichprobengröße von  $n = 71,887$  verringert sich die korrigierte Effektgröße auf  $\rho = .01$ . Unabhängig von der Exklusion bzw. Inklusion der Studie von Reeve et al. (2006) kann ein Populationseffekt von Null nicht ausgeschlossen werden. Da die erklärte Varianz deutlich kleiner als 75% ist, kann ein mittlerer Effekt verschiedener Subpopulationen nicht ausgeschlossen werden.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durchgeführt wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = .10$ . Nach der anschließenden Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) verringert sich die korrigierte Effektgröße von  $\rho = .10$  auf  $\rho = .02$ . Ohne die Exklusion der Studie von Reeve et al. (2006) kann von einem positiven Populationszusammenhang ausgegangen werden, da das Glaubwürdigkeitsintervall die Null nicht mit einschließt. Nach der Exklusion der Studien von Reeve et al. (2006) kann hingegen ein Nulleffekt in der Population nicht mehr ausgeschlossen werden. Auch in dieser Analyse-Variante beträgt die erklärte Varianz deutlich weniger als 75%, weshalb ein mittlerer Effekt heterogener Subpopulationen nicht ausgeschlossen werden kann.

Hinsichtlich des Publikationsbias kann in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien durchgeführt wurde, eine symmetrische Verteilung der publizierten Effektgröße um die mittlere Effektgröße von  $\bar{r} = .02$  mit  $R_0 < 1$  angenommen und ein Publikationsbias ausgeschlossen werden. Aufgrund der Höhe des zweiten Schätzers mit  $L_0 = 7.06$  kann hingegen ein Publikationsbias nicht ausgeschlossen werden. Nach der durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens, kann ein Publikationsbias sowohl für den positiven als auch negativen Wertebereich publizierter Effektgrößen nicht ausgeschlossen werden. Um ein Publikationsbias ausschließen zu können, müssten unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich weitere 152 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) und unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich weitere 120 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, kann aufgrund der Höhe der Schätzer mit  $R_0 = 1$  und  $L_0 =$

1.41 von einer asymmetrischen Verteilung der publizierten Effektgröße um die mittlere Effektgröße ausgegangen werden. Dementsprechend sollte ein Publikationsbias nicht ausgeschlossen werden. Nach der durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens, kann auch in dieser Analyse-Variante jeweils ein Publikationsbias für den positiven und negativen Wertebereich publizierter Effektgrößen nicht ausgeschlossen werden. Um ein Publikationsbias ausschließen zu können, müssten unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich weitere 234 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) und unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich weitere 192 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein.

Mit keinem Moderator konnten mehr als 72.21% der Varianz erklärt werden. Aufgrund der Ergebnisse ist davon auszugehen, dass weitere Moderatoren zum Zusammenhang zwischen  $g$  und  $V$  existieren, die allerdings im Rahmen der vorliegenden Arbeit nicht identifiziert wurden.

### 5.11 Ergebnisse zur Fluiden Intelligenz und Verträglichkeit

*Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien:*

In der Variante beträgt die inkludierte Studienanzahl  $k = 22$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .02$  ( $SD = .00$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 1,215$  (Range:  $n = 64$  bis  $20,175$ ;  $SD = 4,272.17$ ;  $N = 26,719$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. –überschätzung korrigierte Effektgröße beträgt  $\rho = .02$  ( $SD_{\rho} = .00$ ). Das Glaubwürdigkeitsintervall schließt die Null nicht mit ein und die erklärte Varianz beträgt 100%. (Tabelle 56).

Nach der anschließenden Exklusion der Studie von De Fruyt et al. (2006) mit einer Stichprobengröße von  $n = 20,175$  verringert sich die unkorrigierte Effektgröße von  $\bar{r}_{bare-bones} = .02$  ( $SD = .00$ ) auf  $\bar{r}_{bare-bones} = .00$  ( $SD = .00$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 311$  (Range:  $n = 64$  bis  $2,658$ ;  $SD = 563.17$ ;  $N = 6,544$ ). Die artefaktkorrigierte Effektgröße verringert sich hierbei von  $\rho = .02$  ( $SD_{\rho} = .00$ ) auf  $\rho = .01$  ( $SD_{\rho} = .00$ ). Die erklärte Varianz beträgt 100% (Tabelle 56).

Unabhängig von der Exklusion bzw. Inklusion der Studie von De Fruyt et al. (2006,  $n = 20,175$ ) kann von einem positiven, wenn auch sehr kleinen Effekt in der Population ausgegangen werden, da das Glaubwürdigkeitsintervall die Null nicht mit einschließt. Überdies kann die Höhe der

artefaktkorrigierten Effektgröße von  $\rho = .02$  bzw.  $\rho = .01$  aufgrund der erklärten Varianz von 100% generalisiert werden. Moderatoranalysen wurden in dem Fall nicht durchgeführt.

**Tabelle 56**

Metaanalytische Ergebnisse zum Zusammenhang zwischen Gf und V

| Gf/Verträglichkeit                   | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | %Var.  | 90% CV | $\bar{r}$ | $SD_r$ | %Var   |
|--------------------------------------|----------|----------|--------|-------------|--------|--------|-----------|--------|--------|
| kodiert nach Autor(en) <sup>1</sup>  | 26,719   | 22       | .02    | .000        | 100.00 | .02    | .02       | .000   | 100.00 |
| kodiert nach Autor(en) <sup>1a</sup> | 6,544    | 21       | .01    | .000        | 100.00 | .01    | .00       | .000   | 100.00 |
| kodiert nach CHC <sup>2</sup>        | 36,662   | 42       | .00    | .022        | 79.71  | -.02   | .01       | .017   | 79.78  |
| kodiert nach CHC <sup>2a</sup>       | 16,487   | 41       | .00    | .019        | 89.36  | -.03   | -.01      | .015   | 91.99  |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und -überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; %Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, in denen eine Zuordnung von Gf explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von Gf explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von De Fruyt et al. (2006, *n* = 20,175) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde und die Studie von De Fruyt et al. (2006, *n* = 20,175) exkludiert ist.

*Zuordnung der Fähigkeitskonstrukte gemäß CHC-Theorie:*

In dieser Analyse-Variante beträgt die inkludierte Studienzahl *k* = 42 bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .01$  ( $SD = .02$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 873$  (Range: *n* = 64 bis 20,175;  $SD = 3,116.58$ ; *N* = 36,662). Die artefaktkorrigierte Effektgröße beträgt  $\rho = .00$  ( $SD_{\rho} = .02$ ). Die erklärte Varianz beträgt 79.71%.

Nach der anschließenden Exklusion der Studie von De Fruyt et al. (2006, *n* = 71,887) beträgt die unkorrigierte Effektgröße  $\bar{r}_{bare-bones} = -.01$  ( $SD = .02$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 402$  (Range: *n* = 64 bis 2,982;  $SD = 637.59$ ; *N* = 16,487). Die artefaktkorrigierte Effektgröße beträgt  $\rho = .00$  ( $SD_{\rho} = .02$ ). Die erklärte Varianz beträgt 89.36% (Tabelle 56). Unabhängig von der Exklusion bzw. Inklusion der Studie von De Fruyt et al. (2006) kann in dieser Analyse-Variante von einem Nulleffekt in der Population ausgegangen werden. Überdies werden mehr als 75% der Varianz erklärt, weshalb auch in dieser Analyse-Variante keine Moderatoranalysen durchgeführt wurden.

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe des Schätzers mit  $R_0 < 1$  von einer symmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .00$  ausgegangen und ein Publikationsbias ausgeschlossen werden. Der zweite Schätzer zeigt mit  $L_0 = 1.37$  eine asymmetrische Verteilung der publizierten Effektgrößen an, weshalb in dem Fall ein Publikationsbias nicht ausgeschlossen werden kann. Nach der Exklusion von zwei Effektgrößen ( $L_0^+ = 2$ ) mit negativen Vorzeichen und den größten Beträgen kann eine symmetrische Verteilung hinsichtlich der Summe der Ränge publizierter Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .01$  angenommen und ein Publikationsbias ausgeschlossen werden. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 21 Stichprobengrößen real publizierter Studien und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 57 aufgeführt.

**Tabelle 57**

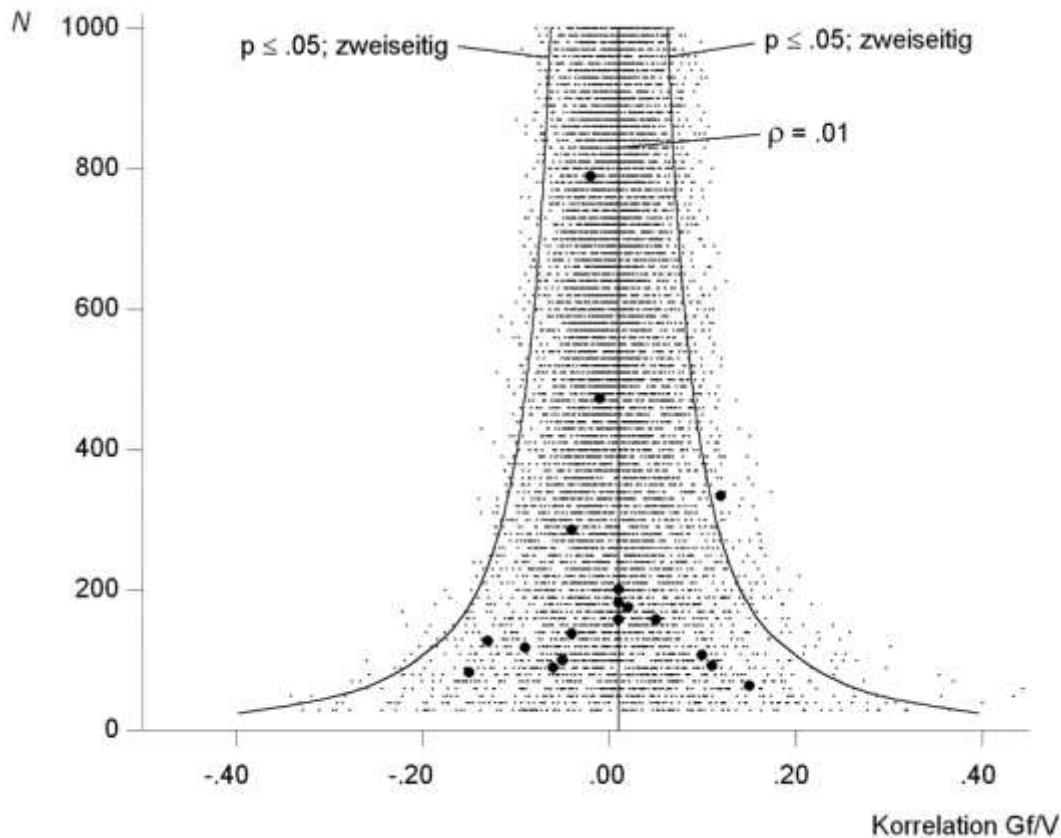
Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | <i>N</i>      | <i>k</i> (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | <i>k</i> (Simulation) <sup>3</sup> |
|-----------|---------------|-----------------------------------|-----------------------|------------------------------------|
| 1.        | 60 – 100      | 7                                 | 33.33                 | 1,667                              |
| 2.        | 110 – 150     | 4                                 | 19.05                 | 953                                |
| 3.        | 160 – 200     | 5                                 | 23.81                 | 1,191                              |
| 4.        | 260 – 300     | 1                                 | 4.76                  | 238                                |
| 5.        | 310 – 350     | 1                                 | 4.76                  | 238                                |
| 6.        | 460 – 500     | 1                                 | 4.76                  | 238                                |
| 7.        | 760 – 800     | 1                                 | 4.76                  | 238                                |
| 8.        | 2,660 – 2,700 | 1                                 | 4.76                  | 238                                |
| $\Sigma$  |               | <b>21</b>                         | <b>100.00</b>         | <b>5,006<sup>4</sup></b>           |

Anmerkungen.  $p = .00$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten *k* für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt *k* von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt *k* = 5,006.

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index beträgt für den positiven Wertebereich .03 und hat sich aus einem Verhältnis von 154 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,852 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Hinsichtlich der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich hat sich ein Index aus der Simulation von .02 aus einem Verhältnis von 92 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,914 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet (Abbildung 34). Da in dem Fall die Indizes der Simulation kleiner sind als der extremste mögliche Wert des empirischen

Indexes .05 (1/21), werden die Indizes aus der Simulation auf .05 aufgerundet (siehe Kapitel 4.6.7).



**Abbildung 34.**  $\rho = .01$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 57) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktionen im positiven und negativen Wertebereich mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien.

Ein Publikationsbias ist für den positiven Wertebereich auszuschließen, da der Index aus den tatsächlich publizierten Effektgrößen mit dem Referenzwert aus der Simulation übereinstimmt. Für den negativen Wertebereich sollte ebenfalls ein Publikationsbias ausgeschlossen werden, da sich bereits bei einer signifikanten publizierten Effektgröße ( $p \leq .05$ , zweiseitig) der Index von .00 auf .05 erhöhen würde und nicht ausgeschlossen werden kann, dass eine signifikante Effektgröße mit negativen Vorzeichen zum Zusammenhang zwischen Gf und V existiert.

Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durchgeführt wurde:

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe der Schätzer mit  $R_0 = 1.5$  von einer asymmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.01$  ausgegangen werden. Ein Publikationsbias sollte dementsprechend nicht ausgeschlossen werden. Durch die Exklusion von zwei Effektgrößen ( $R_0^+ = 2$ ) mit positiven Vorzeichen und den höchsten Beträgen kann eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.02$  angenommen und ein Publikationsbias ausgeschlossen werden. Der zweite Schätzer zeigt mit  $L_0 = 1.46$  ebenfalls eine asymmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.01$  an. Nach der Exklusion von zwei Effektgrößen ( $L_0^+ = 2$ ), allerdings mit negativen Vorzeichen und den höchsten Beträgen, kann von einer Symmetrie hinsichtlich der Summe aller Ränge (siehe A4) publizierter Effektgrößen um die mittlere Effektgrößen von  $\bar{r} = .00$  ausgegangen und dementsprechend ein Publikationsbias ausgeschlossen werden. Darüber hinaus wurde auch in dieser Analyse-Variante eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 41 Stichprobengrößen real publizierter Studien und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 58 aufgeführt.

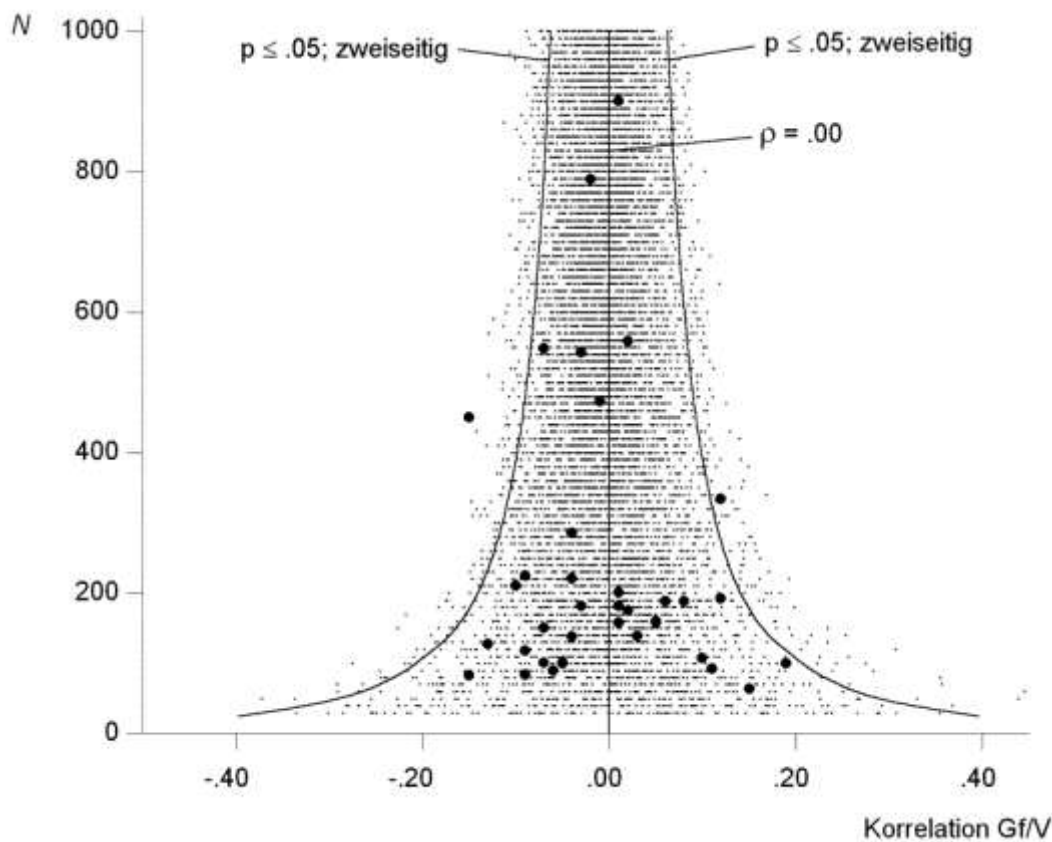
**Tabelle 58**

Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation

| Intervall | $N$           | $k$ (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | $k$ (Simulation) <sup>3</sup> |
|-----------|---------------|------------------------------|-----------------------|-------------------------------|
| 1.        | 60 – 100      | 10                           | 24.39                 | 1,220                         |
| 2.        | 110 – 150     | 6                            | 14.63                 | 732                           |
| 3.        | 160 – 200     | 10                           | 24.39                 | 1,220                         |
| 4.        | 210 – 250     | 3                            | 7.32                  | 366                           |
| 5.        | 260 – 300     | 1                            | 2.44                  | 122                           |
| 6.        | 310 – 350     | 1                            | 2.44                  | 122                           |
| 7.        | 410 – 450     | 1                            | 2.44                  | 122                           |
| 8.        | 460 – 500     | 1                            | 2.44                  | 122                           |
| 9.        | 510 – 550     | 2                            | 4.88                  | 244                           |
| 10.       | 560 – 600     | 1                            | 2.44                  | 122                           |
| 11.       | 760 – 800     | 1                            | 2.44                  | 122                           |
| 12.       | 860 – 900     | 1                            | 2.44                  | 122                           |
| 13.       | 1,810 – 1,850 | 1                            | 2.44                  | 122                           |
| 14.       | 2,660 – 2,700 | 1                            | 2.44                  | 122                           |
| 15.       | 2,960 – 3,000 | 1                            | 2.44                  | 122                           |
| $\Sigma$  |               | <b>41</b>                    | <b>100.00</b>         | <b>5,002<sup>4</sup></b>      |

Anmerkungen.  $p = .01$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten  $k$  für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt  $k$  von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt  $k = 5,002$ .

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .02 hat sich aus einem Verhältnis von 116 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,886 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Da die artefaktkorrigierte Effektgröße in dieser Analyse-Variante  $\rho = .00$  beträgt, gilt der Referenzindex von .02 sowohl für die Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen als auch positiven Wertebereich. Der aus den tatsächlich publizierten Effektgrößen berechnete Index beträgt sowohl für den positiven als auch negativen Wertebereich .02 und unterscheidet sich damit nicht von dem Index aus der Simulation. Ein Publikationsbias ist demnach auszuschließen (Abbildung 35).



**Abbildung 35.**  $\rho = .00$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 58) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktionen im positiven und negativen Wertebereich mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie (McGrew, 2009).

*Zusammenfassung zum Zusammenhang zwischen der Fluiden Intelligenz und Verträglichkeit:*

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien vorgenommen wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = .02$ . Nach der anschließenden Exklusion der Studie von De Fruyt et al. (2006;  $n = 20,175$ ) beträgt die



korrigierte Effektgröße  $\rho = .01$ . Unabhängig von der Exklusion bzw. Inklusion der Studie von De Fruyt et al. (2006) kann in dieser Analyse-Variante von einem positiven, wenn auch sehr kleinen Effekt, in der Population ausgegangen werden.

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = .00$ . Unabhängig von der Exklusion bzw. Inklusion von De Fruyt et al. (2006) kann in dieser Analyse-Variante von einem Nulleffekt in der Population ausgegangen werden.

Hinsichtlich eines Publikationsbias kann in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde, ein Publikationsbias nicht ausgeschlossen werden ( $L_0 = 1.37$ ). Nach der durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens kann ein Publikationsbias sowohl für den positiven als auch negativen Wertebereich ausgeschlossen werden.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, kann aufgrund der Höhe des Schätzers mit  $R_0 = 1.5$  und  $L_0 = 1.46$  ein Publikationsbias nicht ausgeschlossen werden. Allerdings fiel dabei auf, dass die „trim“-Schritte für  $R_0$  im positiven Wertebereich und für  $L_0 = 1.46$  im negativen Wertebereich erforderlich waren, um eine symmetrische Verteilung der Effektgrößen zu generieren. Nach der durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens kann sowohl für den positiven als auch negativen Wertebereich ein Publikationsbias ausgeschlossen werden.

## **5.12 Ergebnisse zur Kristallisierten Intelligenz und Verträglichkeit**

*Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien:*

In der Variante beträgt die inkludierte Studienanzahl  $k = 8$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .13$  ( $SD = .03$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 9,460$  (Range:  $n = 90$  bis  $71,887$ ;  $SD = 25,230.71$ ;  $N = 75,680$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. –überschätzung korrigierte Effektgröße beträgt  $\rho = .18$  ( $SD_\rho = .05$ ). Das Glaubwürdigkeitsintervall schließt die Null nicht mit ein und die erklärte Varianz beträgt 8.75%. (Tabelle 59).

Nach der anschließenden Exklusion der Studie von Reeve et al. (2006) mit einer Stichprobengröße von  $n = 71,887$ , verringert sich die unkorrigierte Effektgröße von  $\bar{r}_{bare-bones} = .13$  ( $SD = .03$ ) auf

## Kapitel 5 – Ergebnisse

$\bar{r}_{bare-bones} = -.02$  ( $SD = .00$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 542$  (Range:  $n = 90$  bis 1,821;  $SD = 613.60$ ;  $N = 3,793$ ). Die artefaktkorrigierte Effektgröße verringert sich hierbei von  $\rho = .18$  ( $SD_{\rho} = .05$ ) auf  $\rho = -.02$  ( $SD_{\rho} = .00$ ). Das Glaubwürdigkeitsintervall schließt die Null nicht mit ein und die erklärte Varianz beträgt 100% (Tabelle 59).

Bleibt die Studie von Reeve et al. (2006) inkludiert, kann von einem positiven Effekt in der Population ausgegangen werden, da das Glaubwürdigkeitsintervall die Null nicht mit einschließt. Allerdings sollte die Höhe des artefaktkorrigierten Effektes nicht generalisiert werden, da die erklärte Varianz lediglich 8.75% beträgt und somit nicht auszuschließen ist, dass ein mittlerer Effekt heterogener Subpopulationen vorliegt. Nach der Exklusion der Studie von Reeve et al. (2006) beträgt die erklärte Varianz 100%. Offenbar wird der Zusammenhang zwischen Gc und V von einer einzelnen Studie mit einer sehr großen Stichprobengröße moderiert. Außerdem kann nach der Exklusion von Reeve et al. (2006) von einem negativen, wenn auch sehr kleinen Effekt in der Population ausgegangen werden, da die Null nicht im Glaubwürdigkeitsintervall liegt.

**Tabelle 59**

Metaanalytische Ergebnisse zum Zusammenhang zwischen Gc und V

| Gc/Verträglichkeit                   | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | % Var. | 90% CV | $\bar{r}$ | $SD_r$ | % Var  |
|--------------------------------------|----------|----------|--------|-------------|--------|--------|-----------|--------|--------|
| kodiert nach Autor(en) <sup>1</sup>  | 75,680   | 8        | .18    | .045        | 8.75   | .13    | .13       | .033   | 8.41   |
| kodiert nach Autor(en) <sup>1a</sup> | 3,793    | 7        | -.02   | .000        | 100.00 | -.02   | -.02      | .000   | 100.00 |
| kodiert nach CHC <sup>2</sup>        | 77,215   | 21       | .18    | .058        | 12.81  | .10    | .13       | .037   | 15.83  |
| kodiert nach CHC <sup>2a</sup>       | 5,328    | 20       | .01    | .070        | 51.36  | -.08   | .00       | .060   | 51.33  |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und -überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; % Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, in denen eine Zuordnung von Gc explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von Gc explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von Reeve et al. (2006,  $n = 71,887$ ) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde und die Studie von Reeve et al. (2006,  $n = 71,887$ ) exkludiert ist.

### Zuordnung der Fähigkeitskonstrukte gemäß CHC-Theorie:

In dieser Analyse-Variante beträgt die inkludierte Studienzahl  $k = 21$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .13$  ( $SD = .04$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 3,677$  (Range:  $n = 92$  bis 71,887;  $SD = 15,630.09$ ;  $N = 77,215$ ). Die artefaktkorrigierte Effektgröße beträgt  $\rho = .18$  ( $SD_{\rho} = .06$ ). Die erklärte Varianz beträgt 12.81% (Tabelle 59).

Nach der Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) beträgt die unkorrigierte Effektgröße  $\bar{r}_{bare-bones} = .00$  ( $SD = .06$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 266$  (Range:  $n = 92$  bis  $789$ ;  $SD = 198.17$ ;  $N = 5,328$ ). Die artefaktkorrigierte Effektgröße verringert sich von  $\rho = .18$  ( $SD_{\rho} = .06$ ) auf  $\rho = .01$  ( $SD_{\rho} = .07$ ). Das Glaubwürdigkeitsintervall schließt die Null mit ein und die erklärte Varianz beträgt 51.36% (Tabelle 59).

Durch die Inklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) kann von einem positiven Effekt in der Population ausgegangen werden, da die Null nicht im Glaubwürdigkeitsintervall liegt. Durch die erklärte Varianz, die deutlich kleiner als 75% ist, kann nicht ausgeschlossen werden, dass ein Effekt heterogener Subpopulationen vorliegt. Nach der Exklusion der Studie von Reeve et al. (2006) kann ein Nulleffekt in der Population nicht mehr ausgeschlossen werden. Die erklärte Varianz nimmt dabei um 38.55% zu. Dennoch kann auch nach der Exklusion der Studie von Reeve et al. (2006) ein Effekt heterogener Subpopulationen nicht ausgeschlossen werden. Dementsprechend wurden in dieser Analyse-Variante Moderatoranalysen gerechnet (siehe unten).

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe der Schätzer mit  $R_0 < 1$  und  $L_0 < 1$  von einer symmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.02$  ausgegangen werden. Demnach kann ein Publikationsbias ausgeschlossen werden. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der sieben Stichprobengrößen real publizierter Studien und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 60 aufgeführt.

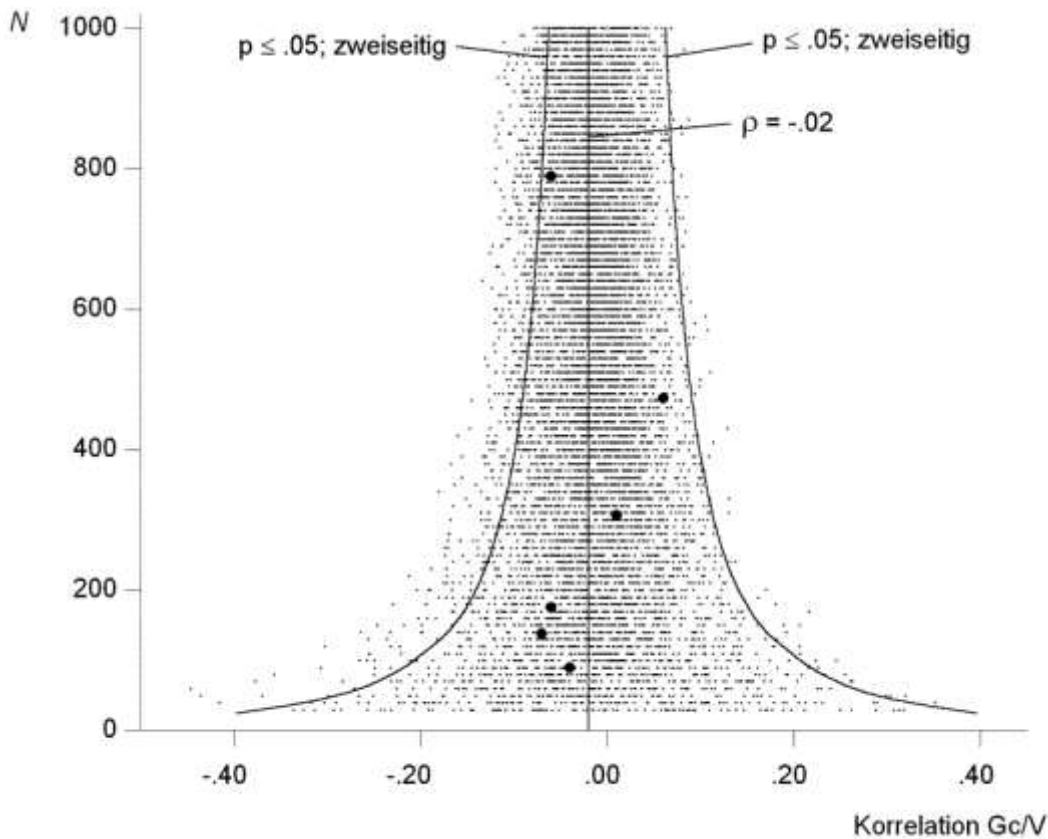
**Tabelle 60**

Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation.

| Intervall | <i>N</i>      | <i>k</i> (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | <i>k</i> (Simulation) <sup>3</sup> |
|-----------|---------------|-----------------------------------|-----------------------|------------------------------------|
| 1.        | 60 – 100      | 1                                 | 14.29                 | 715                                |
| 2.        | 110 – 150     | 1                                 | 14.29                 | 715                                |
| 3.        | 160 – 200     | 1                                 | 14.29                 | 715                                |
| 4.        | 310 – 350     | 1                                 | 14.29                 | 715                                |
| 5.        | 460 – 500     | 1                                 | 14.29                 | 715                                |
| 6.        | 760 – 800     | 1                                 | 14.29                 | 715                                |
| 7.        | 1,810 – 1,850 | 1                                 | 14.29                 | 715                                |
| $\Sigma$  |               | <b>7</b>                          | <b>100.00</b>         | <b>5,005<sup>4</sup></b>           |

Anmerkungen.  $p = .02$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten *k* für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt *k* von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt *k* = 5,005.

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .02 hat sich aus einem Verhältnis von 119 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,886 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich berechnet. Hinsichtlich der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich hat sich ein Index aus der Simulation von .02 aus einem Verhältnis von 91 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,914 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet (Abbildung 36). Da in dem Fall die Indizes der Simulation kleiner sind als der extremste mögliche Wert des empirischen Indexes .14 (1/7), werden die Indizes aus der Simulation auf .14 aufgerundet (siehe Kapitel 4.6.7).



**Abbildung 36.**  $\rho = -.02$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 60) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktionen im positiven und negativen Wertebereich mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien.

Bereits bei einer zusätzlichen publizierten Effektgröße, die signifikant ( $p \leq .05$ , zweiseitig) wäre, würde sich der empirische Index von  $.00$  auf  $.14$  erhöhen. Da nicht ausgeschlossen werden kann, dass eine signifikante Effektgröße zum Zusammenhang zwischen Gc und V existiert, kann ein Publikationsbias sowohl für den positiven als auch negativen Wertebereich ausgeschlossen werden.

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durchgeführt wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe der Schätzer mit  $R_0 = 1$  von einer asymmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .02$  ausgegangen und ein Publikationsbias nicht ausgeschlossen werden. Durch die Exklusion einer Effektgröße mit positiven Vorzeichen und den größten Betrag kann eine symmetrische Verteilung der publizierten

Effektgrößen um die mittlere, neu bestimmte Effektgröße von  $\bar{r} = .01$  mit  $R_0 < 1$  angenommen und ein Publikationsbias ausgeschlossen werden. Der zweite Schätzer zeigt mit  $L_0 < 1$  eine Symmetrie hinsichtlich der summierten Ränge publizierter Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .02$  an. Demnach kann in dem Fall ein Publikationsbias ausgeschlossen werden. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 20 Stichprobengrößen real publizierter Studien und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 61 aufgeführt.

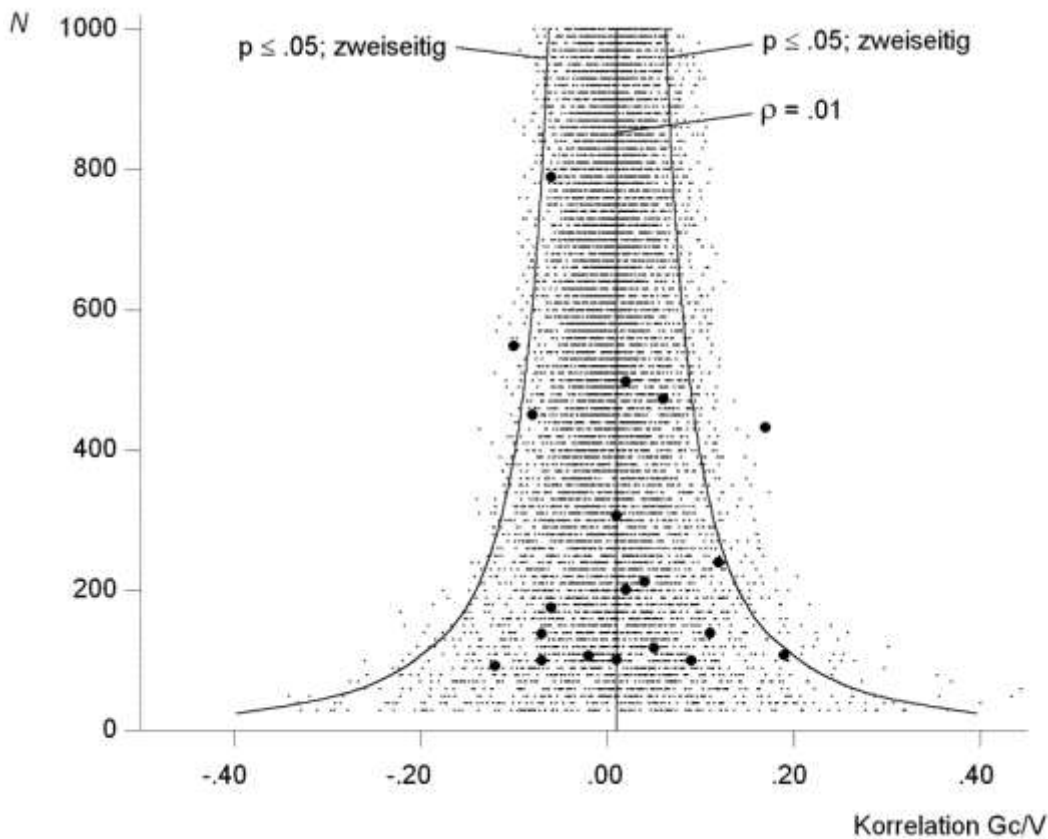
**Tabelle 61**

Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation.

| Intervall | $N$       | $k$ (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | $k$ (Simulation) <sup>3</sup> |
|-----------|-----------|------------------------------|-----------------------|-------------------------------|
| 1.        | 60 – 100  | 4                            | 20.00                 | 1,000                         |
| 2.        | 110 – 150 | 5                            | 25.00                 | 1,250                         |
| 3.        | 160 – 200 | 2                            | 10.00                 | 500                           |
| 4.        | 210 – 250 | 2                            | 10.00                 | 500                           |
| 5.        | 310 – 350 | 1                            | 5.00                  | 250                           |
| 6.        | 410 – 450 | 2                            | 10.00                 | 500                           |
| 7.        | 460 – 500 | 2                            | 10.00                 | 500                           |
| 8.        | 510 – 550 | 1                            | 5.00                  | 250                           |
| 9.        | 760 – 800 | 1                            | 5.00                  | 250                           |
| $\Sigma$  |           | <b>20</b>                    | <b>100.00</b>         | <b>5,000</b>                  |

Anmerkungen.  $p = .01$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten  $k$  für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt  $k$  von 5,000.

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .03 unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich hat sich aus einem Verhältnis von 158 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,842 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Hinsichtlich der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich hat sich ein Index aus der Simulation von .02 aus einem Verhältnis von 86 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,914 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Auch in dieser Analyse-Variante sind die Indizes der Simulation kleiner als der extremste mögliche Wert des empirischen Indexes von .05 (1/20). Deshalb werden die Indizes aus der Simulation auf .05 aufgerundet (siehe Kapitel 4.6.7). Ein Publikationsbias ist demnach sowohl für den positiven als auch negativen Wertebereich auszuschließen, da die Indizes aus den tatsächlich publizierten Effektgrößen mit den Indizes aus der Simulation übereinstimmen (Abbildung 37).



**Abbildung 37.**  $\rho = .01$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 61) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktionen im positiven und negativen Wertebereich mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie (McGrew, 2009).

*Moderatoranalysen:*

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde, wird der Zusammenhang zwischen Gc und V vollständig durch die Studie von Reeve et al. (2006) mit einer Stichprobengröße von  $n = 71,887$  moderiert. Nach der Exklusion von Reeve et al. (2006) nimmt die erklärte Varianz um 91.25% zu und beträgt 100%. Dementsprechend wurden in der Analyse-Variante keine Moderatoranalysen durchgeführt. In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, nimmt die erklärte Varianz nach der Exklusion von Reeve et al. (2006) ebenfalls deutlich zu (Tabelle 62). Allerdings beträgt die erklärte Varianz auch nach der Exklusion von Reeve et al. (2006) weniger als 75%. Dementsprechend wurden in dieser Analyse-Variante Moderatoranalysen durchgeführt. Die kogLT der publizierten Studien berücksichtigend, wurden die Studien, in denen Gc mit dem GKT erfasst wurde, mit den Studien verglichen, in denen andere kogLT als der GKT eingesetzt wurde.

Darüber hinaus wurden Moderatoranalysen hinsichtlich der eingesetzten Persönlichkeitsinventare durchgeführt. Dabei wurden folgende Persönlichkeitsinventare verglichen:

- NEO-Inventare (NEO-PI-(R) oder NEO-FFI) vs.
- NEO-PI-(R) vs.
- NEO-FFI vs.
- FFM-Inventare, allerdings keine NEO-Inventare.

In der Tabelle 62 sind die metaanalytischen Kennwerte der entsprechenden Moderatoranalysen zum Zusammenhang zwischen Gc und V, bei denen  $k \geq 5$  ist, aufgeführt.

**Tabelle 62**  
Moderatoranalysen zum Zusammenhang zwischen Gc und V

| Gc/E         | Moderator  | N     | k  | $\rho$ | $SD_{\rho}$ | % Var. | 90% CV | $\bar{r}$ | $SD_r$ | % Var  |
|--------------|------------|-------|----|--------|-------------|--------|--------|-----------|--------|--------|
| <u>kogLT</u> |            |       |    |        |             |        |        |           |        |        |
| CHC          | GKT        | 832   | 6  | .04    | .000        | 100.00 | .04    | .03       | .000   | 100.00 |
| CHC          | ohne GKT   | 4,496 | 14 | .01    | .080        | 41.83  | -.10   | .00       | .066   | 42.12  |
| <u>PSK</u>   |            |       |    |        |             |        |        |           |        |        |
| CHC          | NEO        | 4,273 | 17 | .01    | .073        | 48.04  | -.08   | .01       | .066   | 48.22  |
| CHC          | NEO-PI-(R) | 1,603 | 8  | .04    | .000        | 100.00 | .04    | .04       | .000   | 100.00 |
| CHC          | NEO-FFI    | 2,670 | 9  | -.02   | .110        | 33.91  | .12    | .00       | .082   | 33.65  |

*Anmerkungen.* N = Stichprobengröße; k = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und –überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; % Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); CHC = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde; kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; GKT = General Knowledge Test; die Studien inkludiert, die den GKT zur Erfassung der kristallisierten Intelligenzausprägung verwendet haben; ohne GKT = die Studien inkludiert, die zur Erfassung der kristallisierten Intelligenz andere kogLT als den GKT verwendet haben; NEO = die Studien inkludiert, in denen der NEO-PI-(R) und der NEO-FFI berücksichtigt wurden; NEO-PI-(R) = die Studien inkludiert, in denen der NEO-PI-R und der NEO-PI verwendet wurde; NEO-FFI = die Studien berücksichtigt, in denen der NEO-FFI eingesetzt wurde; grau markiert sind die Zeilen, in denen mehr als 75% der Varianz erklärt wird.

In den Studien, in denen der GKT zur Erfassung von Gc eingesetzt wurde, werden 100% der Varianz erklärt (Tabelle 62, grau markiert). Dabei kann von einem Zusammenhang von  $\rho = .04$  für diese Subpopulation ausgegangen werden. Offenbar sind die konzeptionellen Unterschiede zwischen dem GKT und anderen kogLT in Bezug auf Gc relevant für die Zusammenhänge mit V. Darüber hinaus wird in den Studien, in denen der NEO-PI-(R) zur Erfassung von V verwendet wurde, ebenfalls 100% der Varianz erklärt (Tabelle 62, grau markiert). Der artefaktkorrigierte Effekt für diese Subpopulation beträgt  $\rho = .04$ . Die konzeptionellen Unterschiede zwischen dem



## *Kapitel 5 – Ergebnisse*

NEO-PI-(R) und Nicht-NEO-PI-(R)-Inventaren scheinen in Bezug auf V relevant für die Zusammenhänge mit Gc zu sein.

### *Zusammenfassung zum Zusammenhang zwischen der Kristallisierten Intelligenz und Verträglichkeit:*

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien vorgenommen wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = .18$ . Hierbei kann von einem positiven Effekt in der Population ausgegangen werden, da das Glaubwürdigkeitsintervall die Null nicht mit einschließt. Allerdings sollte die Höhe des positiven Populationszusammenhangs nicht generalisiert werden, da erklärte Varianz kleiner als 75% ist und dementsprechend ein Effekt heterogener Subpopulationen nicht ausgeschlossen werden kann. Nach der Exklusion der Studie von Reeve et al. (2006) mit einer Stichprobengröße von  $n = 71,887$ ) beträgt die erklärte Varianz 100%. Die artefaktkorrigierte Effektgröße von  $\rho = -.02$  kann generalisiert werden.

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) erfolgte, beträgt die artefaktkorrigierte Effektgröße  $\rho = .18$ . Die Höhe der artefaktkorrigierten Effektgröße von  $\rho = .18$  sollte nicht generalisiert werden, da die erklärte Varianz deutlich kleiner als 75% ist. Allerdings kann von einem positiver Effekt in der Population ausgegangen werden, da das Glaubwürdigkeitsintervall die Null nicht mit einschließt. Nach der Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) verringert sich die artefaktkorrigierte Effektgröße auf  $\rho = .01$ . Ein Nulleffekt ist dabei nicht auszuschließen. Aufgrund der erklärten Varianz, die kleiner als 75% ist, wurden in dieser Analyse-Variante Moderatoranalysen durchgeführt.

Hinsichtlich eines Publikationsbias kann in der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien vorgenommen wurde, von einer symmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.02$  mit  $R_0 < 1$  und  $L_0 < 1$  ausgegangen werden. Dementsprechend kann ein Publikationsbias ausgeschlossen werden. Nach der durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens kann ebenfalls ein Publikationsbias ausgeschlossen werden. In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) erfolgte, beträgt der erste Schätzer  $R_0 = 1$ . Dementsprechend ist ein Publikationsbias, aufgrund einer asymmetrischen Verteilung publizierter Effektgrößen, nicht

auszuschließen. Der zweite Schätzer zeigt hingegen mit  $L_0 < 1$  eine Symmetrie hinsichtlich der Summe aller Ränge publizierter Effektgrößen an, weshalb ein Publikationsbias ausgeschlossen werden kann. Nach der durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens kann ein Publikationsbias im positiven und negativen Wertebereich ausgeschlossen werden.

Moderatoranalysen wurden ausschließlich in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, durchgeführt. In den Studien, in denen der GKT oder der NEO-PI-(R) verwendet wurden, wurde jeweils 100% der Varianz erklärt. Die artefaktkorrigierte Effektgröße beträgt jeweils  $\rho = .04$ . Dementsprechend kann von einem positiven, kleinen Effekt für diese Subpopulationen ausgegangen werden.

### 5.13 Ergebnisse zur Allgemeinen Intelligenz und Gewissenhaftigkeit

*Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien:*

In der Variante beträgt die inkludierte Studienanzahl  $k = 77$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .06$  ( $SD = .06$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 1,178$  (Range:  $n = 10$  bis  $71,887$ ;  $SD = 8,167.74$ ;  $N = 90,777$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. -überschätzung korrigierte Effektgröße beträgt  $\rho = .07$  ( $SD_\rho = .09$ ). Das Glaubwürdigkeitsintervall schließt die Null mit ein und die erklärte Varianz beträgt 16.81%. (Tabelle 63).

Nach der anschließenden Exklusion der Studie von Reeve et al. (2006) mit einer Stichprobengröße von  $n = 71,887$  verringert sich die unkorrigierte Effektgröße von  $\bar{r}_{bare-bones} = .06$  ( $SD = .06$ ) auf  $\bar{r}_{bare-bones} = -.02$  ( $SD = .09$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 247$  (Range:  $n = 10$  bis  $1,886$ ;  $SD = 245.70$ ;  $N = 18,890$ ). Die artefaktkorrigierte Effektgröße verringert sich durch die Exklusion der Studie von Reeve et al. (2006) von  $\rho = .07$  ( $SD_\rho = .09$ ) auf  $\rho = -.04$  ( $SD_\rho = .10$ ). Das Glaubwürdigkeitsintervall schließt die Null mit ein und die erklärte Varianz beträgt 34.57%. (Tabelle 63).

Unabhängig von der Exklusion bzw. Inklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) kann in dieser Analyse-Variante ein Nulleffekt in der Population zum Zusammenhang zwischen  $g$  und  $G$  nicht ausgeschlossen werden. Da die erklärte Varianz deutlich kleiner als 75% ist, wurden Moderatoranalysen durchgeführt (siehe unten).

**Tabelle 63**

Metaanalytische Ergebnisse zum Zusammenhang zwischen g und G

| g/Gewissenhaftigkeit                 | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | %Var. | 90% CV | $\bar{r}$ | $SD_r$ | %Var  |
|--------------------------------------|----------|----------|--------|-------------|-------|--------|-----------|--------|-------|
| kodiert nach Autor(en) <sup>1</sup>  | 90,777   | 77       | .07    | .085        | 16.81 | -.04   | .06       | .059   | 19.70 |
| kodiert nach Autor(en) <sup>1a</sup> | 18,890   | 76       | -.04   | .100        | 34.57 | .09    | -.02      | .090   | 33.33 |
| kodiert nach CHC <sup>2</sup>        | 90,289   | 75       | .08    | .085        | 15.33 | -.03   | .07       | .066   | 15.93 |
| kodiert nach CHC <sup>2a</sup>       | 18,402   | 74       | .02    | .128        | 19.07 | -.14   | .01       | .132   | 18.86 |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und -überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; %Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist.

*Zuordnung der Fähigkeitskonstrukte gemäß CHC-Theorie:*

In dieser Analyse-Variante beträgt die inkludierte Studienzahl *k* = 75 bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .07$  ( $SD = .07$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 1,204$  (Range: *n* = 38 bis 71,887;  $SD = 8273,92$ ; *N* = 90,289). Die artefaktkorrigierte Effektgröße beträgt  $\rho = .08$  ( $SD_{\rho} = .09$ ). Das Glaubwürdigkeitsintervall schließt die Null auch in dieser Analyse-Variante mit ein und die erklärte Varianz beträgt 15.33% (Tabelle 63).

Nach der anschließenden Exklusion der Studie von Reeve et al. (2006, *n* = 71,887) beträgt die unkorrigierte Effektgröße  $\bar{r}_{bare-bones} = .01$  ( $SD = .13$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 249$  (Range: *n* = 38 bis 900;  $SD = 175.40$ ; *N* = 18,402). Die artefaktkorrigierte Effektgröße verringert sich durch die Exklusion von Reeve et al. (2006) von  $\rho = .08$  ( $SD_{\rho} = .09$ ) auf  $\rho = .02$  ( $SD_{\rho} = .13$ ). Das Glaubwürdigkeitsintervall schließt die Null mit ein und die erklärte Varianz beträgt 19.07% (Tabelle 63).

Auch in dieser Analyse-Variante kann ein Nulleffekt in der Population nicht ausgeschlossen werden. Aufgrund der erklärten Varianz, die kleiner als 75% ist, wurden Moderatoranalysen durchgeführt (siehe unten).

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe der Schätzer mit  $R_0 < 1$  und  $L_0 < 1$  von einer symmetrischen Verteilung

der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.01$  ausgegangen werden. Dementsprechend ist ein Publikationsbias auszuschließen. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 76 Stichprobengrößen real publizierter Studien und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 64 aufgeführt.

**Tabelle 64**

Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation.

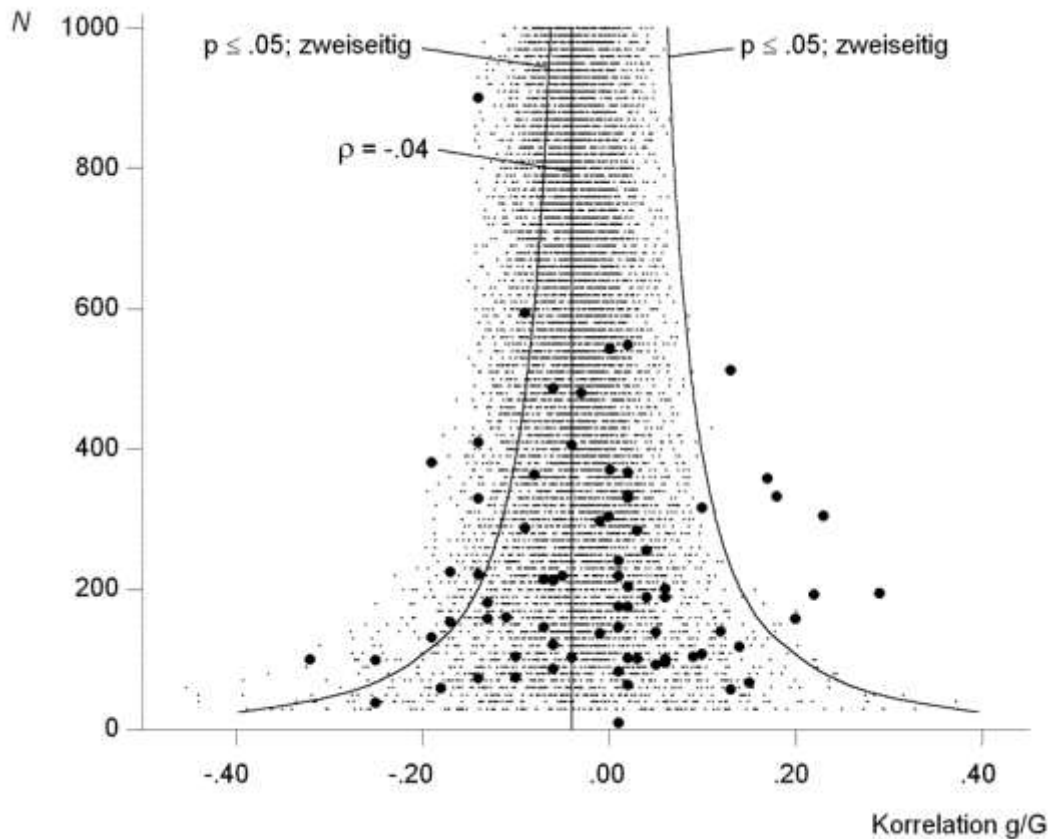
| Intervall | <i>N</i>      | <i>k</i> (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | <i>k</i> (Simulation) <sup>3</sup> |
|-----------|---------------|-----------------------------------|-----------------------|------------------------------------|
| 1.        | 10 – 50       | 2                                 | 2.63                  | 132                                |
| 2.        | 60 – 100      | 18                                | 23.68                 | 1,184                              |
| 3.        | 110 – 150     | 11                                | 14.47                 | 724                                |
| 4.        | 160 – 200     | 12                                | 15.79                 | 790                                |
| 5.        | 210 – 250     | 7                                 | 9.21                  | 461                                |
| 6.        | 260 – 300     | 5                                 | 6.58                  | 329                                |
| 7.        | 310 – 350     | 6                                 | 7.89                  | 395                                |
| 8.        | 360 – 400     | 5                                 | 6.58                  | 329                                |
| 9.        | 410 – 450     | 2                                 | 2.63                  | 132                                |
| 10.       | 460 – 500     | 2                                 | 2.63                  | 132                                |
| 11.       | 510 – 550     | 3                                 | 3.95                  | 198                                |
| 12.       | 560 – 600     | 1                                 | 1.32                  | 66                                 |
| 13.       | 860 – 900     | 1                                 | 1.32                  | 66                                 |
| 14.       | 1,860 – 1,900 | 1                                 | 1.32                  | 66                                 |
| $\Sigma$  |               | <b>76</b>                         | <b>100.00</b>         | <b>5,004<sup>4</sup></b>           |

Anmerkungen.  $p = -.04$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten *k* für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt *k* von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt *k* = 5,004.

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .08 unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich hat sich aus einem Verhältnis von 381 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,623 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Hinsichtlich der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich hat sich ein Index von .01 aus einem Verhältnis von 32 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,972 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet.

Der aus den tatsächlich publizierten Effektgrößen berechnete Index unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich beträgt .16 (12 signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 64 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 36). Bezüglich der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich

hat sich ein Index aus den tatsächlich publizierten Effektgrößen von .09 berechnet (sieben signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 69 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 38).



**Abbildung 38.**  $\rho = -.04$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 64) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktionen im positiven und negativen Wertebereich mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien.

Hinsichtlich der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich sollte ein Publikationsbias nicht ausgeschlossen werden, da weitere 67 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von 12 signifikanten ( $p \leq .05$ , zweiseitig) zu 131 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein müssten, um einen Index von .08 – ohne Publikationsbias – zu erreichen. Ebenso sollte unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich ein Publikationsbias nicht ausgeschlossen werden. Um hierbei einen Index von .01 – ohne Publikationsbias – zu erreichen, müssten weitere 424 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von sieben signifikanten ( $p \leq .05$ , zweiseitig) zu 493 nichtsignifikanten Effektgrößen ( $p > .05$ ,

zweiseitig) publiziert sein. Aufgrund der Ergebnisse sollte ein Publikationsbias sowohl im positiven als auch negativen Wertebereich nicht ausgeschlossen werden (Abbildung 38). Offenbar sind mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde.

Anhand der Abbildung 36 kann der Eindruck entstehen, dass die artefaktkorrigierte Effektgröße von  $\rho = -.04$  die Grundgesamtheit der tatsächlich publizierten Studien mit Stichprobengrößen von  $n < 1000$  nicht optimal abbildet. Es scheint, dass der Betrag der korrigierten Effektgröße von  $\rho = -.04$  zu groß ist (Abbildung 38). Insgesamt sind weniger Effektgrößen links der artefaktkorrigierten Effektgröße von  $\rho = -.04$  mit 29 von insgesamt 76 Effektgrößen publiziert. Diese asymmetrische Verteilung der publizierten Effektgrößen um die korrigierte Effektgröße von  $\rho = -.04$  ist teilweise auf Artefaktkorrekturen, möglicherweise auf die Gewichtung der Stichprobengrößen, zurückzuführen (siehe Kapitel 4.4). Um Veränderungen der artefaktkorrigierten Effektgrößen nach der Exklusion von Studien mit großen Stichprobengrößen einschätzen zu können, wurde eine weitere Analyse, ohne die Studie von Boudreau, Boswell, Judge und Bretz (2001) mit einer Effektgröße von  $r = -.13$  und einer Stichprobengröße von  $n = 1,886$  durchgeführt. Nach der Exklusion dieser Studie verringert sich der Betrag der artefaktkorrigierten Effektgröße von  $\rho = -.04$  auf  $\rho = -.02$ . In einer gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführten Simulation beträgt der Index bei  $\rho = -.02$  für den negativen Wertebereich .04 (212 signifikante,  $p \leq .05$ , zweiseitig; 4,792 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig) und für den positiven Wertebereich .01 (47 signifikante,  $p \leq .05$ , zweiseitig; 4,957 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig). Unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich müssten, um einen Index von .04 – ohne Publikationsbias – zu erreichen, weitere 175 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von 11 signifikanten ( $p \leq .05$ , zweiseitig) zu 239 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Hinsichtlich der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich müssten, um einen Index von .01 – ohne Publikationsbias - zu erreichen, weitere 425 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von sieben signifikanten ( $p \leq .05$ , zweiseitig) zu 493 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Dementsprechend kann unabhängig der zusätzlichen Exklusion der Studie von Boudreau et al. (2001) konsistent ein Publikationsbias für den negativen und positiven Wertebereich angenommen werden, bei dem zu wenig nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) im Verhältnis zu signifikanten Effektgröße ( $p \leq .05$ , zweiseitig) zur Publikation gelangt sind.

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durchgeführt wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe des Schätzers mit  $R_0 < 1$  von einer symmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .00$  ausgegangen und ein Publikationsbias ausgeschlossen werden. Aufgrund der Höhe des zweiten Schätzers mit  $L_0 = 1.84$  kann hingegen ein Publikationsbias nicht ausgeschlossen werden. Durch die Exklusion von zwei Effektgrößen mit positiven Vorzeichen und den größten Beträgen kann eine Symmetrie hinsichtlich der Summe aller Ränge publizierter Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.01$  mit  $L_0 < 1$  angenommen und ein Publikationsbias ausgeschlossen werden. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 74 Stichprobengrößen real publizierter Studien und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 65 aufgeführt.

**Tabelle 65**

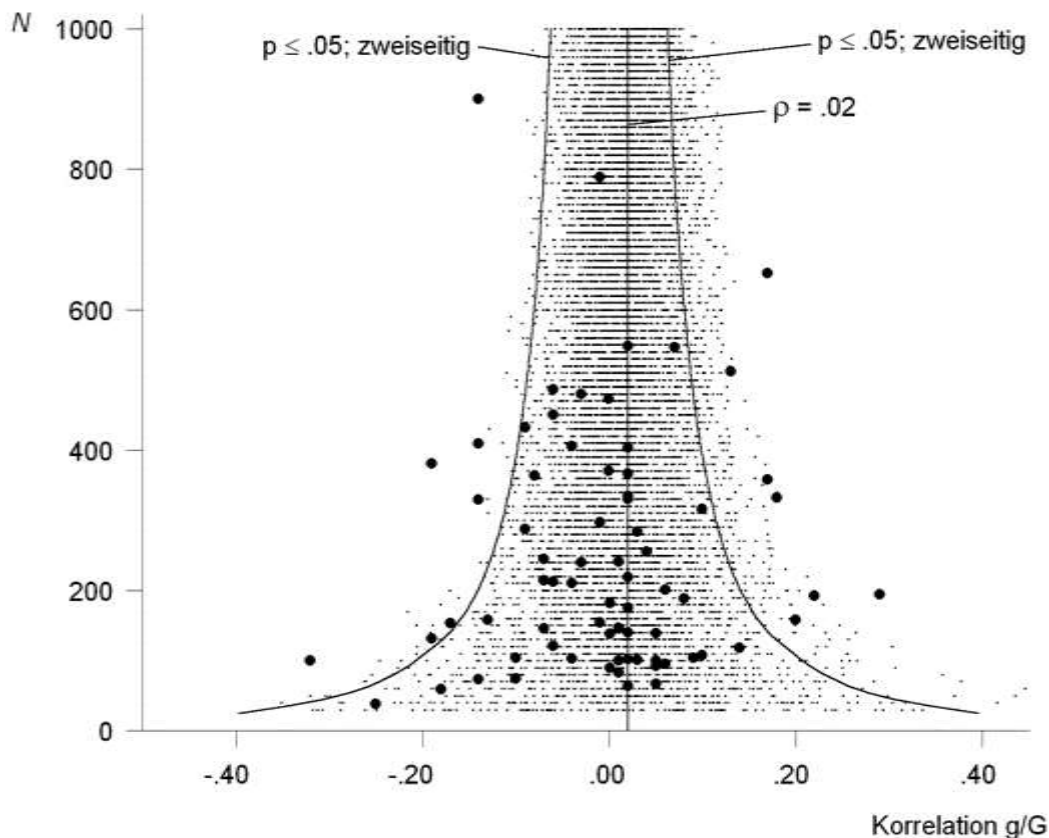
Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation.

| Intervall | <i>N</i>  | <i>k</i> (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | <i>k</i> (Simulation) <sup>3</sup> |
|-----------|-----------|-----------------------------------|-----------------------|------------------------------------|
| 1.        | 10 – 50   | 1                                 | 1.35                  | 68                                 |
| 2.        | 60 – 100  | 18                                | 24.32                 | 1,216                              |
| 3.        | 110 – 150 | 12                                | 16.22                 | 811                                |
| 4.        | 160 – 200 | 8                                 | 10.81                 | 541                                |
| 5.        | 210 – 250 | 7                                 | 9.46                  | 473                                |
| 6.        | 260 – 300 | 4                                 | 5.41                  | 271                                |
| 7.        | 310 – 350 | 5                                 | 6.76                  | 338                                |
| 8.        | 360 – 400 | 6                                 | 8.11                  | 406                                |
| 9.        | 410 – 450 | 4                                 | 5.41                  | 271                                |
| 10.       | 460 – 500 | 3                                 | 4.05                  | 203                                |
| 11.       | 510 – 550 | 3                                 | 4.05                  | 203                                |
| 12.       | 610 – 650 | 1                                 | 1.35                  | 68                                 |
| 13.       | 760 – 800 | 1                                 | 1.35                  | 68                                 |
| 14.       | 860 – 900 | 1                                 | 1.35                  | 68                                 |
| $\Sigma$  |           | <b>74</b>                         | <b>100.00</b>         | <b>5,005<sup>4</sup></b>           |

*Anmerkungen.*  $p = .02$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten *k* für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt *k* von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt *k* = 5,005.

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .05 unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich hat sich aus einem

Verhältnis von 232 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,773 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Hinsichtlich der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich hat sich ein Index von .01 aus einem Verhältnis von 51 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,954 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Der aus den tatsächlich publizierten Effektgrößen berechnete Index beträgt jeweils für den positiven und negativen Wertebereich .09 (sieben signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 67 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 39).



**Abbildung 39.**  $\rho = .02$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 65) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktionen im positiven und negativen Wertebereich mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie (McGrew, 2009).

Um ein Index von .05 – ohne Publikationsbias – zu erreichen, müssten weitere 56 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von sieben signifikanten ( $p \leq .05$ , zweiseitig) zu 123 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Um einen Index von .01 – ohne Publikationsbias – zu erreichen, müssten weitere 426 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von sieben signifikanten ( $p \leq .05$ ,



zweiseitig) zu 493 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Aufgrund der Ergebnisse kann ein Publikationsbias im positiven und negativen Wertebereich nicht ausgeschlossen werden. Dabei ist davon auszugehen, dass mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt sind, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde.

Anhand der Abbildung 39 ist auffällig, dass im linken Bereich und rechten Bereich, deutlich außerhalb der simulierten Effektgrößen, fünf Effektgrößen publiziert sind. In Tabelle 66 sind zu den fünf Effektgrößen studienspezifische Merkmale aufgeführt. Dabei werden in den US-amerikanischen Studien von Judge et al. (1999) sowie Barrick et al. (1998), in denen Untersuchungen im beruflichen Kontext durchgeführt wurden, positive Effektgrößen zwischen  $g$  und  $G$  berichtet. Hingegen werden in den Studien von Furnham (2005) sowie Allik und Realo (1997), in denen Schüler bzw. Studenten untersucht wurden, negative Korrelationen zwischen  $g$  und  $G$  berichtet. Möglicherweise sind die unterschiedlichen Vorzeichen der Effektgrößen beim Zusammenhang zwischen  $g$  und  $G$  darauf zurückzuführen, dass eine Formung der Persönlichkeit im schulischen bzw. studentischen Alltag stattfindet, die sich im beruflichen Alltag mit Zunahme des Alters stabilisiert (Terracciano, McCrae & Costa, 2010; Rindfuss, 1991; Roberts & DelVecchio, 2000) und sich positiv auf die Intelligenz auswirkt (vgl. Stumm et al., 2011; Zeidner, 1995). Dessen ungeachtet sind keine eindeutigen Auffälligkeiten bzw. Gemeinsamkeiten festzustellen, weshalb diese fünf Studien Effektgrößen berichten, die deutlich außerhalb der simulierten Effektgrößen liegen (Tabelle 66).

**Tabelle 66**

Deskriptive Angaben publizierter Studien, in denen die Effektgröße des Zusammenhangs zwischen  $g$  und  $V$  von den gezogenen Effektgrößen aus der adaptierten Simulation abweicht.

| Studie                                      | kogLT             | PSK      | Land    | Alter                                    | Stichprobe                        | Thema   |
|---|-------------------|----------|---------|--|-----------------------------------|---|
| Judge, Higgins, Thoresen und Barrick (1999) | SBT               | CQ-Set   | USA     | $M =$ nicht spezifiziert                 | Erwachsenen ( $n = 194$ )         | Längstschnittstudie zu Persönlichkeit, Intelligenz und Karriere                       |
| Barrick, Stewart, Neubert und Mount (1998)  | WPT               | PCI      | USA     | $M = 41.66$<br>( $SD = 9.59$ )           | Angestellte ( $n = 652$ )         | Auswirkungen von Gruppeneffekten auf Teamprozesse und –effektivität im Arbeitskontext |
| Furnham (2005)                              | WPT               | NEO-PI-R | GB      | $M = 18.17$<br>( $SD = 2.46$ )<br>Range: | Studenten ( $n = 100$ )           | selbsteingeschätzte vs. psychometrische Intelligenz und Persönlichkeit                |
| Allik und Realo (1997)                      | unbekannter kogLT | NEO-PI   | Estland | $M = 19.10$<br>( $SD = 2.30$ )           | angehende Studenten ( $n = 381$ ) | akademische Fähigkeiten   |
| Moutafi, Furnham und Crump (2003)           | WGCTA und GMA     | NEO-PI-R | GB      | $M = 44.00$<br>( $SD = n.b.$ )           | Angestellte ( $n = 900$ )         | Vorhersage von Intelligenz durch Persönlichkeit                                       |

*Anmerkungen.* kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; SBT = Stanford Binet Test; California Q-Set (Block 1971); PCI = Personal Characteristics Inventory (Barrick & Mount, 1995).

*Moderatoranalysen:*

Zum Zusammenhang zwischen  $g$  und  $G$  beträgt die Varianz des artefaktkorrigierten Effektes ausnahmslos weniger als 75%. Dementsprechend wurden Moderatoranalysen durchgeführt. Die kognitiven Leistungstests der publizierten Studien berücksichtigend, wurden die Studien, in denen der WPT zur Erfassung der allgemeinen Intelligenzausprägung eingesetzt wurde, mit den Studien verglichen, die andere kogLT als den WPT verwendet haben. Darüber hinaus wurden Moderatoranalysen hinsichtlich der eingesetzten Persönlichkeitsinventare durchgeführt. Dabei wurden folgende Persönlichkeitsinventare verglichen:

- NEO-Inventare vs.
- NEO-PI-(R) vs.
- NEO-FFI vs.
- FFM-Inventare (keine NEO-Inventare).

In der Tabelle 67 sind die metaanalytischen Kennwerte der entsprechenden Moderatoranalysen zum Zusammenhang zwischen  $g$  und  $G$  aufgeführt.

**Tabelle 67**  
Moderatoranalysen zum Zusammenhang zwischen g und G

| g/G                     | Moderator   | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | % Var. | 90% CV | $\bar{r}$ | $SD_r$ | % Var  |
|-------------------------|-------------|----------|----------|--------|-------------|--------|--------|-----------|--------|--------|
| <i>kogLT</i>            |             |          |          |        |             |        |        |           |        |        |
| Autor(en) <sup>1</sup>  | WPT         | 5,601    | 33       | -.01   | .070        | 64.42  | .08    | -.01      | .056   | 65.63  |
| Autor(en) <sup>1</sup>  | ohne WPT    | 84,847   | 43       | .07    | .082        | 11.50  | -.03   | .06       | .055   | 14.39  |
| Autor(en) <sup>1a</sup> | ohne WPT    | 12,960   | 42       | -.05   | .102        | 26.29  | -.09   | -.03      | .100   | 24.37  |
| CHC <sup>2</sup>        | WPT         | 6,551    | 35       | .03    | .051        | 53.33  | -.04   | .00       | .078   | 46.87  |
| CHC <sup>2</sup>        | ohne WPT    | 84,108   | 41       | .09    | .086        | 10.62  | -.02   | .07       | .062   | 11.07  |
| CHC <sup>2a</sup>       | ohne WPT    | 12,221   | 40       | .01    | .191        | 12.37  | -.23   | .01       | .151   | 12.59  |
| <i>PSK</i>              |             |          |          |        |             |        |        |           |        |        |
| Autor(en) <sup>1</sup>  | FFM, NEO    | 11,912   | 47       | -.05   | .082        | 40.72  | .05    | -.04      | .079   | 39.06  |
| Autor(en) <sup>1</sup>  | NEO-PI-(R)  | 8,124    | 28       | -.07   | .088        | 32.16  | .04    | -.06      | .091   | 29.39  |
| Autor(en) <sup>1</sup>  | NEO-FFI     | 3,788    | 19       | .00    | .000        | 100.00 | .00    | -.01      | .000   | 100.00 |
| Autor(en) <sup>1</sup>  | FFM, no NEO | 6,452    | 27       | -.03   | .096        | 38.93  | .09    | -.01      | .083   | 37.87  |
| CHC <sup>2</sup>        | FFM, NEO    | 11,781   | 47       | -.02   | .060        | 60.82  | .06    | -.02      | .048   | 63.06  |
| CHC <sup>2</sup>        | NEO-PI-(R)  | 7,241    | 27       | -.02   | .081        | 41.85  | .08    | -.02      | .071   | 42.81  |
| CHC <sup>2</sup>        | NEO-FFI     | 4,540    | 20       | -.01   | .000        | 100.00 | -.01   | -.02      | .000   | 100.00 |
| CHC <sup>2</sup>        | FFM, no NEO | 6,095    | 25       | .03    | .065        | 36.26  | -.05   | .01       | .095   | 31.39  |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und -überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; % Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von g explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes g durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist; kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; WPT = Wonderlic Personal Test (hierbei die Studien inkludiert, die den WPT zur Erfassung der allgemeinen Intelligenzausprägung verwendet haben; ohne WPT = die Studien inkludiert, die zur Erfassung der allgemeinen Intelligenz andere kogLT als den WPT verwendet haben; FFM, NEO = die Studien inkludiert, in denen Costa und McCrae-Inventare (NEO-PI, NEO-PI-R, NEO-FFI) eingesetzt wurden; NEO-PI-(R) = die Studien inkludiert, in denen der NEO-PI-R und der NEO-PI verwendet wurde; NEO-FFI = die Studien berücksichtigt, in denen der NEO-FFI eingesetzt wurde; FFM, no NEO = die Studien berücksichtigt, in denen FFM-Inventare, allerdings keine Costa und McCrae Inventare verwendet wurden; grau markiert sind die Zeilen, in denen mehr als 75% der Varianz erklärt wird.

Unabhängig von der Analyse-Variante (Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien vs. CHC-Theorie) wird in den Studien, in denen der NEO-FFI zur Erfassung von G verwendet wurde, 100% der Varianz erklärt (Tabelle 67, grau markiert). Der Unterschied zwischen dem NEO-FFI und Nicht-NEO-FFI-Inventaren führt z.T. zu erheblicher Varianz der

Effektstärken. Offenbar sind die konzeptionellen Unterschiede zwischen dem NEO-FFI und Nicht-NEO-FFI-Inventaren hinsichtlich der Gewissenhaftigkeit in Bezug auf  $g$  relevant.

*Zusammenfassung zum Zusammenhang zwischen der Allgemeinen Intelligenz und Gewissenhaftigkeit:*

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien vorgenommen wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = .07$ . Nach der anschließenden Exklusion der Studie von Reeve et al. (2006) mit einer Stichprobengröße von  $n = 71,887$ ) beträgt die artefaktkorrigierte Effektgröße  $\rho = -.04$ . Unabhängig von der Exklusion bzw. Inklusion der Studie von Reeve et al. (2006) kann ein Nulleffekt in der Population nicht ausgeschlossen werden. Darüber hinaus kann von einem Effekt heterogener Subpopulationen ausgegangen werden, da die erklärte Varianz deutlich kleiner als 75% ist.

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = .08$ . Nach der anschließenden Exklusion der Studie von Reeve et al. (2006) verringert sich die artefaktkorrigierte Effektgröße auf  $\rho = .02$ . Auch in dieser Analyse-Variante kann ein Nulleffekt in der Population nicht ausgeschlossen werden. Aufgrund der erklärten Varianz von maximal 19.07% wurden ebenfalls Moderatoranalysen durchgeführt (Tabelle 63).

Hinsichtlich eines Publikationsbias kann in der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien vorgenommen wurde, von einer symmetrischen Verteilung der publizierten Effektgrößen mit  $R_0 < 1$  und  $L_0 < 1$  ausgegangen und dementsprechend ein Publikationsbias ausgeschlossen werden. Nach den durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens kann hingegen ein Publikationsbias für den positiven und negativen Wertebereich nicht ausgeschlossen werden. Um ein Publikationsbias ausschließen zu können, müssten weitere 67 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich publiziert sein. Unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich müssten weitere 424 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Es ist davon auszugehen, dass mehr signifikante ( $p \leq .05$ , zweiseitig) als nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt sind.

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) erfolgte, kann auf Grund der Höhe der Schätzer mit  $R_0 < 1$  eine symmetrische Verteilung der publizierten Effektgrößen angenommen und ein Publikationsbias ausgeschlossen werden. Aufgrund der Höhe des zweiten Schätzers mit  $L_0 = 1.84$  kann hingegen ein Publikationsbias nicht ausgeschlossen werden. Nach der durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens ist von einem Publikationsbias im positiven und negativen Wertebereich auszugehen, da mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) zur Publikation gelangt sind als nichtsignifikante Effektgröße ( $p > .05$ , zweiseitig). Um ein Publikationsbias ausschließen zu können, müssten weitere 56 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich publiziert sein. Unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich müssten sogar weitere 426 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein.

In den Studien, in denen der NEO-FFI zur Erfassung von G eingesetzt wurde, konnten unabhängig von der Analyse-Variante (Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien vs. CHC-Theorie) 100% der Varianz erklärt werden. Offenbar sind die konzeptionellen Unterschiede zwischen dem NEO-FFI und Nicht-NEO-FFI-Inventaren hinsichtlich der Gewissenhaftigkeit in Bezug auf g relevant.

#### **5.14 Ergebnisse zur Fluiden Intelligenz und Gewissenhaftigkeit**

*Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien:*

In der Variante beträgt die inkludierte Studienanzahl  $k = 23$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .00$  ( $SD = .05$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 1,168$  (Range:  $n = 64$  bis  $20,175$ ;  $SD = 4,179.80$ ;  $N = 26,873$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. -überschätzung korrigierte Effektgröße beträgt  $\rho = -.02$  ( $SD_\rho = .06$ ). Das Glaubwürdigkeitsintervall schließt die Null mit ein und die erklärte Varianz beträgt 28.13%. (Tabelle 68).

Nach der anschließenden Exklusion der Studie von De Fruyt et al. (2006) mit einer Stichprobengröße von  $n = 20,175$  verringert sich die unkorrigierte Effektgröße von  $\bar{r}_{bare-bones} = .00$  ( $SD = .05$ ) auf  $\bar{r}_{bare-bones} = -.07$  ( $SD = .06$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 304$  (Range:  $n = 64$  bis  $2,658$ ;  $SD = 550.62$ ;  $N = 6,690$ ). Die artefaktkorrigierte Effektgröße verringert

sich hierbei von  $\rho = -.02$  ( $SD_\rho = .06$ ) auf  $\rho = -.07$  ( $SD_\rho = .05$ ). Die Null liegt nicht im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 57.10% (Tabelle 68).

Ohne die Exklusion der Studie von De Fruyt et al. (2006) kann ein Nulleffekt in der Population nicht ausgeschlossen werden, da das Glaubwürdigkeitsintervall die Null mit einschließt. Nach der anschließenden Exklusion der Studie von De Fruyt et al. (2006,  $n = 20,175$ ) verändert sich der Befund substantiell: Es kann von einem negativen Effekt in der Population ausgegangen werden. Unabhängig von der Exklusion bzw. Inklusion der Studie von De Fruyt et al. (2006) wird in dieser Analyse-Variante weniger als 75% der Varianz erklärt, weshalb Moderatoranalysen durchgeführt wurden (siehe unten).

**Tabelle 68**  
Metaanalytische Ergebnisse zum Zusammenhang zwischen Gf und G

| Gf/Gewissenhaftigkeit                | $N$    | $k$ | $\rho$ | $SD_\rho$ | % Var. | 90% CV | $\bar{r}$ | $SD_r$ | % Var |
|--------------------------------------|--------|-----|--------|-----------|--------|--------|-----------|--------|-------|
| kodiert nach Autor(en) <sup>1</sup>  | 26,873 | 23  | -.02   | .060      | 28.13  | .06    | .00       | .049   | 26.52 |
| kodiert nach Autor(en) <sup>1a</sup> | 6,690  | 22  | -.07   | .047      | 57.10  | -.01   | -.07      | .060   | 47.94 |
| kodiert nach CHC <sup>2</sup>        | 36,846 | 45  | -.04   | .074      | 26.03  | .06    | -.02      | .062   | 24.09 |
| kodiert nach CHC <sup>2a</sup>       | 16,663 | 44  | -.08   | .065      | 40.74  | .00    | -.08      | .059   | 43.09 |

*Anmerkungen.*  $N$  = Stichprobengröße;  $k$  = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und -überschätzung korrigierte Effektgröße;  $SD_\rho$  = Standardabweichung der artefaktkorrigierten Effektgröße; % Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, in denen eine Zuordnung von Gf explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von Gf explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von De Fruyt et al. (2006,  $n = 20,175$ ) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde und die Studie von De Fruyt et al. (2006,  $n = 20,175$ ) exkludiert ist.

*Zuordnung der Fähigkeitskonstrukte gemäß CHC-Theorie:*

In dieser Analyse-Variante beträgt die inkludierte Studienzahl  $k = 45$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = -.02$  ( $SD = .06$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 819$  (Range:  $n = 64$  bis  $20,175$ ;  $SD = 3,015.31$ ;  $N = 36,846$ ). Die artefaktkorrigierte Effektgröße beträgt  $\rho = -.04$  ( $SD_\rho = .07$ ). Das Glaubwürdigkeitsintervall schließt die Null mit ein und die erklärte Varianz beträgt 26.03% (Tabelle 68).

Nach der anschließenden Exklusion der Studie von De Fruyt et al. (2006,  $n = 20,175$ ) beträgt die unkorrigierte Effektgröße  $\bar{r}_{bare-bones} = -.08$  ( $SD = .06$ ) bei einer mittleren Stichprobengröße von  $\bar{n}$

## Kapitel 5 – Ergebnisse

= 379 (Range:  $n = 64$  bis 2,982;  $SD = 620.56$ ;  $N = 16,663$ ). Die artefaktkorrigierte Effektgröße verringert sich durch die Exklusion der Studie von Reeve et al. (2006) von  $\rho = -.04$  ( $SD_{\rho} = .07$ ) auf  $\rho = -.08$  ( $SD_{\rho} = .07$ ). Das Glaubwürdigkeitsintervall schließt die Null mit ein und die erklärte Varianz beträgt 40.74% (Tabelle 68).

Unabhängig von der Exklusion bzw. Inklusion der Studie von De Fruyt et al. (2006) kann in dieser Analyse-Variante ein Nulleffekt in der Population nicht ausgeschlossen werden. Darüber hinaus wurden auch in dieser Analyse-Variante Moderatoranalysen durchgeführt, da die erklärte Varianz weniger als 75% beträgt.

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien vorgenommen wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe der Schätzer mit  $R_0 < 1$  und  $L_0 < 1$  von einer symmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.06$  ausgegangen werden. Demnach ist ein Publikationsbias auszuschließen. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 21 Stichprobengrößen real publizierter Studien und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 69 aufgeführt.

**Tabelle 69**

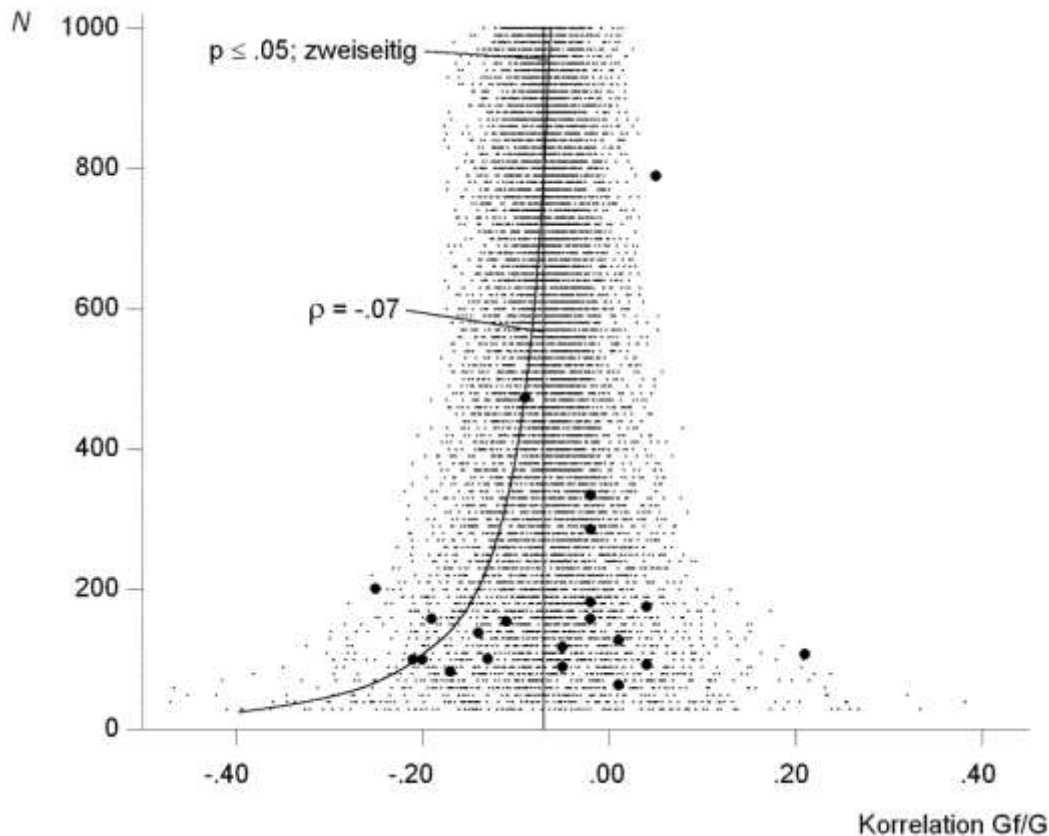
Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation.

| Intervall | $N$       | $k$ (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | $k$ (Simulation) <sup>3</sup> |
|-----------|-----------|------------------------------|-----------------------|-------------------------------|
| 1.        | 60 – 100  | 7                            | 33.33                 | 1,667                         |
| 2.        | 110 – 150 | 5                            | 23.81                 | 1,191                         |
| 3.        | 160 – 200 | 5                            | 23.81                 | 1.191                         |
| 4.        | 260 – 300 | 1                            | 4.76                  | 238                           |
| 5.        | 310 – 350 | 1                            | 4.76                  | 238                           |
| 6.        | 460 – 500 | 1                            | 4.76                  | 238                           |
| 7.        | 760 – 800 | 1                            | 4.76                  | 238                           |
| $\Sigma$  |           | <b>21</b>                    | <b>100.00</b>         | <b>5,001<sup>4</sup></b>      |

Anmerkungen.  $\rho = -.07$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten  $k$  für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt  $k$  von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt  $k = 5,001$ .

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .12 hat sich aus einem Verhältnis von 606 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,395 nichtsignifikanten

Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Der aus den tatsächlich publizierten Effektgrößen berechnete Index von .19 hat sich aus einem Verhältnis von vier signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) zu 17 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) berechnet (Abbildung 40).



**Abbildung 40.**  $\rho = -.07$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 69) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien.

Um ein Index von .12 – ohne Publikationsbias – zu erreichen, müssten 12 weitere nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von sieben signifikanten ( $p \leq .05$ , zweiseitig) zu 29 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Demzufolge kann ein Publikationsbias nicht ausgeschlossen werden, da mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt sind, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde.



Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durchgeführt wurde:

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe des ersten Schätzers mit  $R_0 = 1.5$  von einer asymmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.06$  ausgegangen und ein Publikationsbias nicht ausgeschlossen werden. Nach der Exklusion der zwei Effektgrößen ( $R_0^+ = 2$ ) mit positiven Vorzeichen und den höchsten Beträgen kann eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.07$  angenommen und ein Publikationsbias ausgeschlossen werden. Der zweite Schätzer zeigt mit  $L_0 < 1$  eine symmetrische Verteilung hinsichtlich der Summe aller Ränge publizierter Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.06$  an. Ein Publikationsbias wäre demnach auszuschließen. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 42 Stichprobengrößen real publizierter Studien und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 70 aufgeführt.

**Tabelle 70**

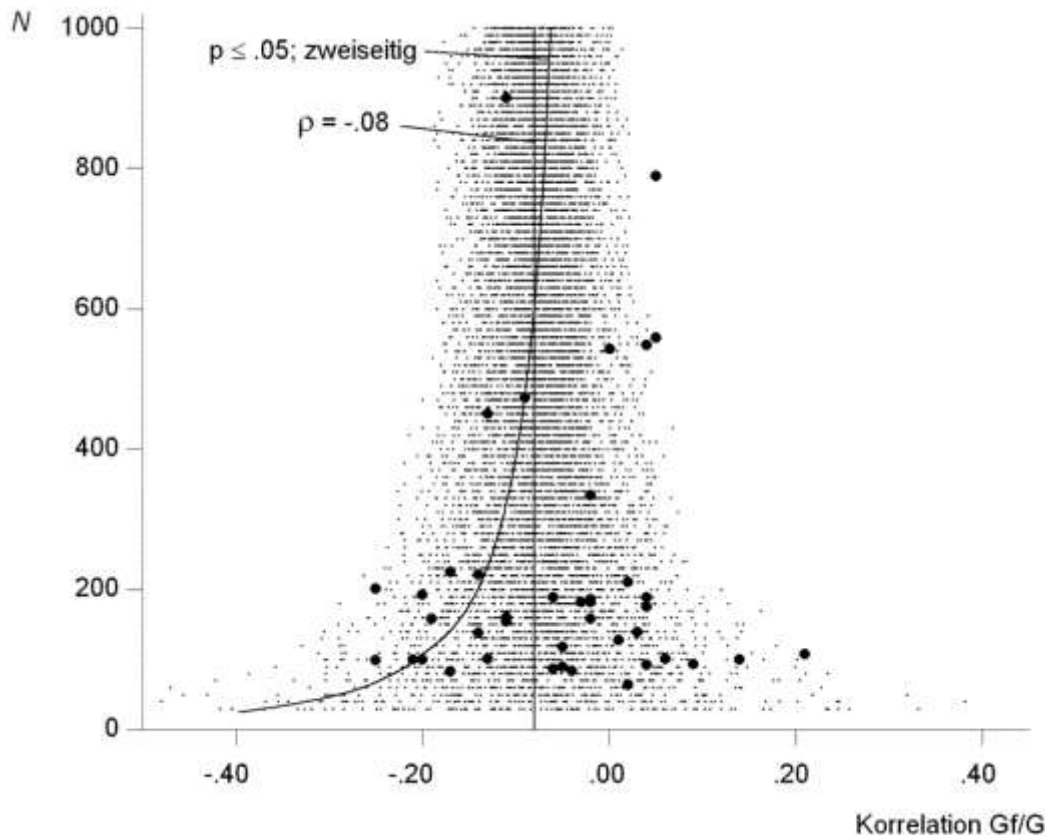
Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation.

| Intervall | <i>N</i>      | <i>k</i> (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | <i>k</i> (Simulation) <sup>3</sup> |
|-----------|---------------|-----------------------------------|-----------------------|------------------------------------|
| 1.        | 60 – 100      | 13                                | 30.95                 | 1,548                              |
| 2.        | 110 – 150     | 6                                 | 14.29                 | 715                                |
| 3.        | 160 – 200     | 11                                | 26.19                 | 1,310                              |
| 4.        | 210 - 250     | 3                                 | 7.14                  | 357                                |
| 5.        | 310 – 350     | 1                                 | 2.38                  | 119                                |
| 6.        | 410 – 450     | 1                                 | 2.38                  | 119                                |
| 7.        | 460 – 500     | 1                                 | 2.38                  | 119                                |
| 8.        | 510 – 550     | 2                                 | 4.76                  | 238                                |
| 9.        | 560 – 600     | 1                                 | 2.38                  | 119                                |
| 10.       | 760 – 800     | 1                                 | 2.38                  | 119                                |
| 11.       | 860 – 900     | 1                                 | 2.38                  | 119                                |
| 12.       | 1,810 – 1,850 | 1                                 | 2.38                  | 119                                |
| <b>Σ</b>  |               | <b>42</b>                         | <b>100.00</b>         | <b>5,001<sup>4</sup></b>           |

Anmerkungen.  $\rho = -.08$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten *k* für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt *k* von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt *k* = 5,001.

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .19 hat sich aus einem Verhältnis von 933 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,068 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Der aus den tatsächlich publizierten Effektgrößen

berechnete Index beträgt .24 (10 signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; 32 nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 41).



**Abbildung 41.**  $\rho = -.08$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 70) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie (McGrew, 2009).

Um ein Index von .19 – ohne Publikationsbias – zu erreichen, müssten 10 weitere nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) in einem Verhältnis von 10 signifikanten ( $p \leq .05$ , zweiseitig) zu 42 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Demzufolge kann auch in dieser Analyse-Variante ein Publikationsbias nicht ausgeschlossen werden. Offenbar sind mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) zur Publikation gelangt, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde.

#### *Moderatoranalysen:*

Die Varianz des artefaktkorrigierten Effektes zum Zusammenhang zwischen Gf und G beträgt ausnahmslos weniger als 75%. Dementsprechend wurden Moderatoranalysen durchgeführt. Die

kognitiven Leistungstests der publizierten Studien berücksichtigend, wurde die Studien, in denen der BRT zur von Gf eingesetzt wurde, mit den Studien verglichen, in denen zur Messung von Gf andere kogLT als der BRT eingesetzt wurden. Darüber hinaus wurden die Studien, in denen Raven Tests zur Erfassung von Gf eingesetzt wurden, mit den Studien verglichen, in denen andere kogLT als die Raven Tests eingesetzt wurden. Darüber hinaus wurden Moderatoranalysen hinsichtlich der eingesetzten Persönlichkeitsinventare durchgeführt. Dabei wurden folgende Persönlichkeitsinventare verglichen:

- NEO-Inventare (NEO-PI-(R) oder NEO-FFI) vs.
- NEO-PI-(R) vs.
- NEO-FFI vs.
- FFM-Inventare, allerdings keine NEO-Inventare.

In Tabelle 71 sind die metaanalytischen Kennwerte der entsprechenden Moderatoranalysen zum Zusammenhang zwischen Gf und G mit  $k \geq 5$  berichtet.

**Tabelle 71**

Moderatoranalysen zum Zusammenhang zwischen Gf und G

| Gf/G                    | Moderator   | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | % Var. | 90% CV | $\bar{r}$ | $SD_r$ | % Var |
|-------------------------|-------------|----------|----------|--------|-------------|--------|--------|-----------|--------|-------|
| <i>kogLT</i>            |             |          |          |        |             |        |        |           |        |       |
| Autor(en) <sup>1</sup>  | BRT         | 1,589    | 11       | -.10   | .100        | 46.09  | .03    | -.09      | .090   | 46.13 |
| Autor(en) <sup>1</sup>  | ohne BRT    | 25,284   | 12       | -.01   | .050        | 24.27  | .05    | .00       | .039   | 23.55 |
| Autor(en) <sup>1a</sup> | ohne BRT    | 5,109    | 11       | -.07   | .022        | 79.10  | -.04   | -.07      | .045   | 51.12 |
| CHC <sup>2</sup>        | BRT         | 1,589    | 11       | -.10   | .100        | 46.09  | .03    | -.09      | .090   | 46.13 |
| CHC <sup>2</sup>        | ohne BRT    | 35,257   | 34       | -.04   | .070        | 23.49  | .05    | -.02      | .059   | 21.73 |
| CHC <sup>2a</sup>       | ohne BRT    | 15,082   | 33       | -.08   | .061        | 39.41  | .00    | -.08      | .055   | 42.25 |
| CHC <sup>2</sup>        | Raven       | 23,242   | 13       | .01    | .047        | 35.56  | -.05   | .01       | .031   | 36.67 |
| CHC <sup>2a</sup>       | Raven       | 3,067    | 12       | -.08   | .056        | 66.38  | -.01   | -.06      | .044   | 67.33 |
| CHC <sup>2</sup>        | ohne Raven  | 13,604   | 32       | -.08   | .067        | 35.64  | .01    | -.08      | .061   | 38.37 |
| <i>PSK</i>              |             |          |          |        |             |        |        |           |        |       |
| Autor(en) <sup>1</sup>  | NEO         | 26,587   | 22       | -.02   | .062        | 26.91  | .06    | .00       | .049   | 25.40 |
| Autor(en) <sup>1a</sup> | NEO         | 6,412    | 21       | -.08   | .048        | 55.86  | -.01   | -.07      | .061   | 46.52 |
| Autor(en) <sup>1</sup>  | NEO-PI-R    | 25,122   | 15       | -.02   | .062        | 20.52  | .06    | .00       | .050   | 19.15 |
| Autor(en) <sup>1a</sup> | NEO-PI-(R)  | 4,947    | 14       | -.08   | .042        | 54.78  | -.03   | -.09      | .052   | 50.41 |
| Autor(en) <sup>1</sup>  | NEO-FFI     | 1,465    | 7        | -.01   | .041        | 84.11  | .04    | .00       | .030   | 84.27 |
| CHC <sup>2</sup>        | NEO         | 35,073   | 39       | -.04   | .073        | 24.45  | .06    | -.02      | .062   | 22.65 |
| CHC <sup>2a</sup>       | NEO         | 14,898   | 38       | -.08   | .065        | 38.80  | .01    | -.08      | .060   | 41.01 |
| CHC <sup>2</sup>        | NEO-PI-(R)  | 32,164   | 26       | -.04   | .075        | 18.39  | .06    | -.02      | .064   | 16.77 |
| CHC <sup>2a</sup>       | NEO-PI-(R)  | 11,989   | 25       | -.09   | .062        | 34.76  | -.01   | -.09      | .053   | 42.35 |
| CHC <sup>2a</sup>       | NEO-FFI     | 2,909    | 13       | -.01   | .045        | 77.71  | .05    | -.01      | .036   | 78.09 |
| CHC <sup>2</sup>        | FFM, no NEO | 1,581    | 5        | -.07   | .029        | 84.03  | -.04   | -.06      | .026   | 81.78 |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und -überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; % Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, in denen eine Zuordnung von Gf explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von Gf explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von De Fruyt et al. (2006, *n* = 20,175) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde und die Studie von De Fruyt et al. (2006, *n* = 20,175) exkludiert ist; kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; BRT = Baddeley Reasoning Test (hierbei die Studien inkludiert, die den BRT zur Erfassung der fluiden Intelligenzausprägung verwendet haben; ohne BRT = die Studien inkludiert, die zur Erfassung der fluiden Intelligenz andere kogLT als den BRT verwendet haben; Raven = die Studien inkludiert, die den BRT zur Erfassung der fluiden Intelligenzausprägung verwendet haben; ohne Raven = die Studien inkludiert, die zur Erfassung der fluiden Intelligenz andere kogLT als den Raven Test verwendet haben; NEO = die Studien inkludiert, in denen ausschließlich Costa und McCrae-Inventare (NEO-PI, NEO-PI-R, NEO-FFI) eingesetzt wurden; NEO-PI-(R) = die Studien inkludiert, in denen der NEO-PI-R und der NEO-PI verwendet wurde; NEO-FFI = die Studien berücksichtigt, in denen der NEO-FFI eingesetzt wurde; FFM, no NEO = die Studien berücksichtigt, in denen FFM-Inventare, allerdings keine Costa und McCrae Inventare verwendet wurden; grau markiert sind die Zeilen, in denen mehr als 75% der Varianz erklärt wird

## *Kapitel 5 – Ergebnisse*

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien vorgenommen wurde, wird in den Studien, in denen andere kogLt als der BRT verwendet wurden, mehr als 75% der Varianz erklärt. Dabei kann eine artefaktkorrigierten Effektgröße von  $\rho = -.07$  für diese Subpopulation angenommen werden. In den Studien, in denen der BRT zur Erfassung von Gf verwendet wurde, werden überraschenderweise nicht mehr als 46.09% der Varianz erklärt. In dem Fall ist davon auszugehen, dass weitere Moderatoren diesen Effekt erklären.

Unabhängig von der Analyse-Variante wird in den Studien, in den der NEO-FFI verwendet wurde, mehr als 75% der Varianz erklärt (Tabelle 71, grau markiert). Dabei kann für diese Subpopulation ein Nulleffekt nicht ausgeschlossen werden, da das Glaubwürdigkeitsintervall die Null mit einschließt. Der Unterschied zwischen dem NEO-FFI und Nicht-NEO-FFI-Inventaren führt zu einer erheblichen Varianz der Effektstärken. Offenbar sind die konzeptionellen Unterschiede zwischen dem NEO-FFI und Nicht-NEO-FFI-Inventaren in Bezug auf G relevant für die Zusammenhänge mit Gf.

Darüber hinaus werden in den Studien, in denen G nicht mit NEO-Inventaren erfasst wurde, mehr als 75% der Varianz erklärt. Dabei kann von einem negativen Zusammenhang von  $\rho = -.07$  ausgegangen werden (Tabelle 71, grau markiert). Offenbar besteht ein relevanter Unterschied zwischen Nicht-NEO-Inventaren und NEO-Inventaren, der sich unterschiedlich auf den Zusammenhang zwischen Gf und G, hinsichtlich eines Nulleffektes, auswirkt.

### *Zusammenfassung zum Zusammenhang zwischen der Fluiden Intelligenz und der Gewissenhaftigkeit:*

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien vorgenommen wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = -.02$ . Nach der Exklusion der Studie von De Fruyt et al. (2006,  $n = 20,175$ ) beträgt die artefaktkorrigierte Effektgröße  $\rho = -.07$ . Durch die Inklusion der Studie von De Fruyt et al. (2006) kann ein Nulleffekt in der Population nicht ausgeschlossen werden. Durch die Exklusion der Studie von De Fruyt et al. (2006) kann hingegen von einem negativen Effekt in der Population ausgegangen werden. Unabhängig von der Exklusion bzw. Inklusion der Studie von De Fruyt (2006) beträgt die erklärte Varianz weniger als 75%.

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = -.04$ . Nach der anschließenden Exklusion der Studie von De Fruyt et al. (2006,  $n = 20,175$ ) beträgt die artefaktkorrigierte Effektgröße  $\rho = -.08$ . Unabhängig von der Exklusion bzw. Inklusion der Studie von De Fruyt et al. (2006) kann in dieser Analyse-Variante ein Nulleffekt in der Population nicht ausgeschlossen werden. Die erklärte Varianz beträgt ebenfalls weniger als 75%.

Hinsichtlich eines Publikationsbias kann in der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien vorgenommen wurde, von einer symmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.06$  mit  $R_0 < 1$  und  $L_0 < 1$  ausgegangen werden. Demnach ist ein Publikationsbias auszuschließen. Nach der durchgeführten Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens kann ein Publikationsbias nicht ausgeschlossen werden, da zu viele signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sind als man bei einer zufälligen Verteilung erwarten würde.

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, kann auf Grund der Höhe des Schätzer mit  $R_0 = 1.5$  von einer asymmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.06$  ausgegangen werden. Ein Publikationsbias kann demnach nicht ausgeschlossen werden. Durch die Exklusion von zwei Effektgrößen ( $R_0^+ = 2$ ) kann mit  $R_0 < 1$  ein Publikationsbias ausgeschlossen werden. Aufgrund der Höhe des zweiten Schätzers kann mit  $L_0 < 1$  ein Publikationsbias ausgeschlossen werden. Nach der durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens kann ein Publikationsbias nicht ausgeschlossen werden, da im Vergleich zu einer zufälligen Verteilung signifikanter ( $p \leq .05$ , zweiseitig) und nichtsignifikanter Effektgrößen ( $p > .05$ , zweiseitig) zu viele signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) publiziert wurden.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien vorgenommen wurde, kann in den Studien, in denen andere kogLt als der BRT zur Erfassung von Gf eingesetzt wurden, von einer artefaktkorrigierten Effektgröße von  $\rho = -.07$  ausgegangen werden. In den Studien, in denen der NEO-FFI zur Erfassung von G verwendet wurde, wurden mehr als 75% der Varianz erklärt. Dabei ist von einem Nulleffekt auszugehen. In den Studien, in denen G mit FFM-Inventaren, ohne NEO-Inventare, erfasst wurden, kann hingegen

von einem negativen Zusammenhang von  $\rho = -.07$  ausgegangen werden. Der Unterschied zwischen NEO-FFI-Inventaren und FFM-Inventaren (ohne NEO-Inventare) führen zu unterschiedlichen Ergebnissen, ob ein Nulleffekt auszuschließen ist oder nicht. Offenbar bestehen hierbei konzeptionelle Unterschiede, die in Bezug auf G relevant für die Zusammenhänge mit Gf sind.

### 5.15 Ergebnisse zur Kristallisierten Intelligenz und Gewissenhaftigkeit

*Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien:*

In der Variante beträgt die inkludierte Studienanzahl  $k = 9$  bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .09$  ( $SD = .03$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 8,437$  (Range:  $n = 90$  bis  $71,887$ ;  $SD = 23,800$ ;  $N = 75,834$ ). Die um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium sowie die Varianzunterschätzung bzw. -überschätzung korrigierte Effektgröße beträgt  $\rho = .13$  ( $SD_{\rho} = .05$ ). Die Null liegt nicht im Glaubwürdigkeitsintervall und die erklärte Varianz beträgt 8.68%. (Tabelle 72).

Nach der anschließenden Exklusion der Studie von Reeve et al. (2006) mit einer Stichprobengröße von  $n = 71,887$  verringert sich die unkorrigierte Effektgröße von  $\bar{r}_{bare-bones} = .09$  ( $SD = .03$ ) auf  $\bar{r}_{bare-bones} = -.05$  ( $SD = .04$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 506$  (Range:  $n = 90$  bis  $1,821$ ;  $SD = 584.97$ ;  $N = 3,947$ ). Die artefaktkorrigierte Effektgröße verringert sich hierbei von  $\rho = .13$  ( $SD_{\rho} = .05$ ) auf  $\rho = -.07$  ( $SD_{\rho} = .06$ ). Das Glaubwürdigkeitsintervall schließt die Null mit ein und die erklärte Varianz beträgt 52.70% (Tabelle 72).

Durch die Inklusion der Studie von Reeve et al. (2006) kann von einem positiven Effekt in der Population ausgegangen werden, da das Glaubwürdigkeitsintervalls die Null nicht mit einschließt. Allerdings kann ein positiver Effekt heterogener Subpopulationen nicht ausgeschlossen werden, da die erklärte Varianz deutlich kleiner als 75% ist. Nach der anschließenden Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) verändert sich der Befund substantiell: Ein Nulleffekt kann in der Population nicht ausgeschlossen werden, da das Glaubwürdigkeitsintervall die Null mit einschließt. Die erklärte Varianz beträgt auch nach der Exklusion der Studie von Reeve et al. (2006) weniger als 75%. Dementsprechend wurden in dieser Analyse-Variante Moderatoranalysen durchgeführt (siehe unten).

**Tabelle 72**

Metaanalytische Ergebnisse zum Zusammenhang zwischen Gc und G

| Gc/Gewissenhaftigkeit                | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | %Var. | 90% CV | $\bar{r}$ | $SD_r$ | %Var  |
|--------------------------------------|----------|----------|--------|-------------|-------|--------|-----------|--------|-------|
| kodiert nach Autor(en) <sup>1</sup>  | 75,834   | 9        | .13    | .048        | 8.68  | .06    | .09       | .034   | 9.12  |
| kodiert nach Autor(en) <sup>1a</sup> | 3,947    | 8        | -.07   | .057        | 52.70 | .09    | -.05      | .044   | 50.69 |
| kodiert nach CHC <sup>2</sup>        | 77,773   | 23       | .12    | .051        | 16.99 | .06    | .09       | .036   | 18.79 |
| kodiert nach CHC <sup>2a</sup>       | 5,886    | 22       | .00    | .089        | 38.94 | -.12   | -.01      | .077   | 38.93 |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und -überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; %Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, in denen eine Zuordnung von Gc explizit von den Autoren der publizierten Studien vorgenommen wurde; <sup>1a</sup> = Datensätze, in denen eine Zuordnung von Gc explizit von den Autoren der publizierten Studien vorgenommen wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist; <sup>2</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gf durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde; <sup>2a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) durchgeführt wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist.

*Zuordnung der Fähigkeitskonstrukte gemäß CHC-Theorie:*

In dieser Analyse-Variante beträgt die inkludierte Studienzahl *k* = 23 bei einer unkorrigierten Effektgröße von  $\bar{r}_{bare-bones} = .09$  ( $SD = .04$ ) und einer mittleren Stichprobengröße von  $\bar{n} = 3,381$  (Range: *n* = 92 bis 71,887;  $SD = 14,913.87$ ; *N* = 77,773). Die artefaktkorrigierte Effektgröße beträgt  $\rho = .12$  ( $SD_{\rho} = .05$ ). Das Glaubwürdigkeitsintervall schließt die Null nicht mit ein und die erklärte Varianz beträgt 16.99% (Tabelle 72).

Nach der anschließenden Exklusion der Studie von Reeve et al. (2006, *n* = 71,887) beträgt die unkorrigierte Effektgröße  $\bar{r}_{bare-bones} = -.01$  ( $SD = .08$ ) bei einer mittleren Stichprobengröße von  $\bar{n} = 268$  (Range: *n* = 92 bis 789;  $SD = 192.44$ ; *N* = 5,886). Die artefaktkorrigierte Effektgröße verringert sich durch die Exklusion von Reeve et al. (2006) von  $\rho = .12$  ( $SD_{\rho} = .05$ ) auf  $\rho = .00$  ( $SD_{\rho} = .09$ ). Die erklärte Varianz beträgt 38.94% (Tabelle 72).

Durch die Inklusion der Studie von Reeve et al. (2006) kann auch in dieser Analyse-Variante von einem positiven Effekt in der Population ausgegangen werden. Nach der Exklusion der Studie von Reeve et al. (2006) ist ebenfalls ein Effekt von Null in der Population nicht auszuschließen. Da die erklärte Varianz, unabhängig von der Exklusion bzw. Inklusion der Studie von Reeve et al. (2006), kleiner als 75% ist, wurden auch in dieser Analyse-Variante Moderatoranalysen durchgeführt (siehe unten).



*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe der Schätzer mit  $R_0 = 1$  von einer asymmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.03$  ausgegangen werden. Ein Publikationsbias sollte demnach nicht ausgeschlossen werden. Nach der Exklusion einer Effektgröße mit positiven Vorzeichen und dem höchsten Betrag kann eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.05$  angenommen und ein Publikationsbias ausgeschlossen werden. Aufgrund der Höhe des zweiten Schätzer mit  $L_0 < 1$  kann ein Publikationsbias um die mittlere Effektgröße von  $\bar{r} = -.03$  ausgeschlossen werden. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der acht Stichprobengrößen real publizierter Studien und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 73 aufgeführt.

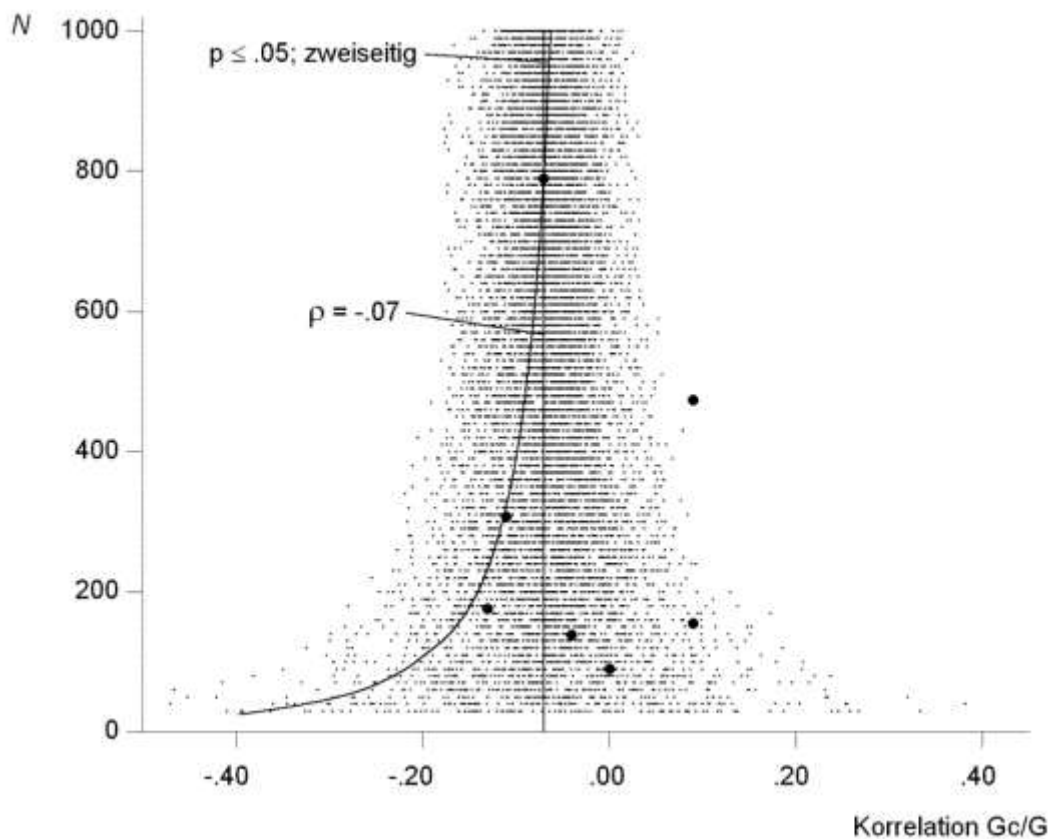
**Tabelle 73**

Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation.

| Intervall | <i>N</i>      | <i>k</i> (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | <i>k</i> (Simulation) <sup>3</sup> |
|-----------|---------------|-----------------------------------|-----------------------|------------------------------------|
| 1.        | 60 – 100      | 1                                 | 12.50                 | 625                                |
| 2.        | 110 – 150     | 2                                 | 25.00                 | 1,250                              |
| 3.        | 160 – 200     | 1                                 | 12.50                 | 625                                |
| 4.        | 310 – 350     | 1                                 | 12.50                 | 625                                |
| 5.        | 560 – 600     | 1                                 | 12.50                 | 625                                |
| 6.        | 760 – 800     | 1                                 | 12.50                 | 625                                |
| 7.        | 1,810 – 1,850 | 1                                 | 12.50                 | 625                                |
| $\Sigma$  |               | <b>8</b>                          | <b>100.00</b>         | <b>5,001<sup>4</sup></b>           |

Anmerkungen.  $p = -.07$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten *k* für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt *k* von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt *k* = 5,001.

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .29 hat sich aus einem Verhältnis von 1,435 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 3,566 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Der aus den tatsächlich publizierten Effektgrößen berechnete Index beträgt .25 (zwei signifikante Effektgrößen,  $p \leq .05$ , zweiseitig; sechs nichtsignifikante Effektgrößen,  $p > .05$ , zweiseitig; Abbildung 42).



**Abbildung 42.**  $\rho = -.07$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 73) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien.

Bereits bei einer weiteren publizierten signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) würde der Index .33 betragen und den Index aus der Simulation mit .29 – ohne Publikationsbias – überschreiten. Dementsprechend kann in dem Fall ein Publikationsbias ausgeschlossen werden.

*Untersuchung eines Publikationsbias in der Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie durchgeführt wurde:*

Nach der „trim and fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Kapitel 4.6.1) kann auf Grund der Höhe des ersten Schätzers mit  $R_0 = 1$  von einer asymmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = -.03$  ausgegangen werden. Ein Publikationsbias ist demnach nicht auszuschließen. Durch die Exklusion einer Effektgröße mit positiven Vorzeichen und dem höchsten Betrag kann eine symmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .00$  mit  $R_0 < 1$  angenommen und ein Publikationsbias ausgeschlossen werden. Hinsichtlich des zweiten Schätzers  $L_0$  kann ein Publikationsbias ebenfalls nicht ausgeschlossen werden. In dem Fall waren zwei „trim“-Schritte

erforderlich, bei denen insgesamt fünf Effektgrößen mit negativen Vorzeichen und den höchsten Beträgen exkludiert wurden. Anschließend konnte eine Symmetrie der publizierter Effektgrößen bei einer mittleren Effektgröße von  $\bar{r} = .05$  angenommen und ein Publikationsbias ausgeschlossen werden. Darüber hinaus wurde eine Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens durchgeführt. Die prozentuale Verteilung der 22 Stichprobengrößen real publizierter Studien und der dementsprechenden Anzahl gezogener Effektgrößen aus der Simulation sind in Tabelle 74 aufgeführt.

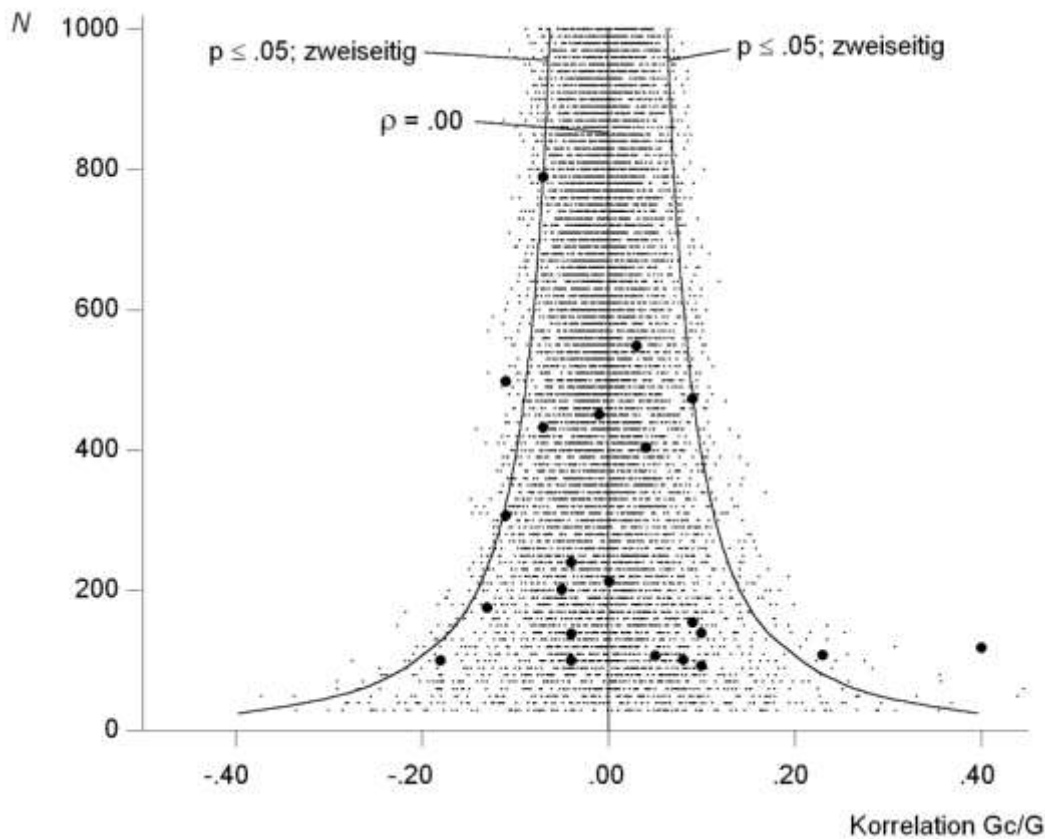
**Tabelle 74**

Zuordnung publizierter Stichproben in Intervalle unter Berücksichtigung der Stichprobengrößen und entsprechend adaptierter Simulation.

| Intervall | $N$       | $k$ (empirisch) <sup>1</sup> | %-Anteil <sup>2</sup> | $k$ (Simulation) <sup>3</sup> |
|-----------|-----------|------------------------------|-----------------------|-------------------------------|
| 1.        | 60 – 100  | 4                            | 18.18                 | 909                           |
| 2.        | 110 – 150 | 6                            | 27.27                 | 1,364                         |
| 3.        | 160 – 200 | 2                            | 9.09                  | 455                           |
| 4.        | 210 – 250 | 2                            | 9.09                  | 455                           |
| 5.        | 310 – 350 | 1                            | 4.55                  | 228                           |
| 6.        | 360 – 400 | 1                            | 4.55                  | 228                           |
| 7.        | 410 – 450 | 2                            | 9.09                  | 455                           |
| 8.        | 460 – 500 | 2                            | 9.09                  | 455                           |
| 9.        | 510 – 550 | 1                            | 4.55                  | 228                           |
| 10        | 760 – 800 | 1                            | 4.55                  | 228                           |
| $\Sigma$  |           | <b>22</b>                    | <b>100.00</b>         | <b>5,005<sup>4</sup></b>      |

Anmerkungen.  $p = .00$ ; <sup>1</sup> = Anzahl publizierter Studien, jeweils nach Stichprobengröße den Intervallen zugeordnet; <sup>2</sup> = prozentualer Anteil des summierten  $k$  für das jeweilige Intervall; <sup>3</sup> = Anzahl gezogener simulierter Effektgrößen je Intervall nach prozentualer Verteilung bei einem gesamt  $k$  von 5,000; <sup>4</sup> = durch vorgenommene Rundungen beträgt  $k = 5,005$ .

Der aus der adaptierten Simulation (Kapitel 4.6.7) gebildete Index von .02 hat sich aus einem Verhältnis von 116 signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) zu 4,889 nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) berechnet. Aufgrund der artefaktkorrigierten Effektgröße von  $\rho = .00$  kann ein Index von .02 sowohl für den positiven als auch negativen Wertebereich als Referenzwert zur Beurteilung eines Publikationsbias angenommen werden (Abbildung 43). Allerdings ist der Index von .02 aus der Simulation kleiner als der extremste mögliche Wert des empirischen Index mit .05 (1/22). Deshalb werden die Indizes aus der Simulation auf .05 aufgerundet (siehe Kapitel 4.6.7).



**Abbildung 43.**  $\rho = .00$ ; tatsächlich publizierte Effektgrößen als schwarze ausgefüllte Kreise abgebildet; aus der Simulation gezogene Effektgrößen als Punkte abgebildet, zu grafischen Zwecken wurden die Prozenanteile der simulierten Effektgrößen (Tabelle 74) nicht berücksichtigt, sondern anhand der korrigierten Effektgröße die Parameter der Simulation 1 verwendet ( $n \geq 30$  bis  $\leq 1,000$ ,  $k = 9,800$ , Kapitel 4.6.4); eingezeichnete Signifikanzfunktion mit  $p \leq .05$  (zweiseitig); Zuordnung der Fähigkeitskonstrukte nach der CHC-Theorie (McGrew, 2009).

Um einen Index von .05 – ohne Publikationsbias – zu erreichen, müssten unter Berücksichtigung der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im negativen Wertebereich weitere 35 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Hinsichtlich der Signifikanzgrenze ( $p \leq .05$ , zweiseitig) im positiven Wertebereich, müssten sogar weitere 53 nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) publiziert sein. Demzufolge kann ein Publikationsbias für den negativen und positiven Wertebereich nicht ausgeschlossen werden. Auf beiden Seiten sind unter Berücksichtigung der Signifikanzgrenzen ( $p \leq .05$ , zweiseitig) mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde.

*Moderatoranalysen:*

Zum Zusammenhang zwischen  $G_c$  und  $G$  beträgt die Varianz des artefaktkorrigierten Effektes weniger als 75%. Dementsprechend wurden Moderatoranalysen durchgeführt. Die kognitiven

Leistungstests der publizierten Studien berücksichtigend, wurden die Studien, in denen ausschließlich der GKT zur Erfassung von Gc eingesetzt wurde mit den Studien verglichen, in denen zur Erfassung von Gc andere kogLT als der GKT eingesetzt wurden. Darüber hinaus wurden Moderatoranalysen hinsichtlich der eingesetzten Persönlichkeitsinventare durchgeführt. Da gewährleistet sein soll, dass die Stichprobenanzahl mindestens  $k \geq 5$  beträgt (siehe Kapitel 4.5), konnten hierbei drei Moderatoren verglichen werden

- NEO-Inventare (NEO-PI-(R) oder NEO-FFI) vs.
- NEO-PI-(R) vs.
- NEO-FFI.

In Tabelle 75 sind die metaanalytischen Kennwerte der entsprechenden Moderatoranalysen zum Zusammenhang zwischen Gc und G aufgeführt.

**Tabelle 75**  
Moderatoranalysen zum Zusammenhang zwischen der Gc und G

| Gc/G              | Moderator    | <i>N</i> | <i>k</i> | $\rho$ | $SD_{\rho}$ | % Var. | 90% CV | $\bar{r}$ | $SD_r$ | % Var |
|-------------------|--------------|----------|----------|--------|-------------|--------|--------|-----------|--------|-------|
|                   | <i>kogLT</i> |          |          |        |             |        |        |           |        |       |
| CHC <sup>1</sup>  | GKT          | 832      | 6        | .09    | .120        | 32.20  | -.06   | .10       | .126   | 31.14 |
| CHC <sup>1</sup>  | ohne GKT     | 76,941   | 17       | .12    | .049        | 14.59  | .06    | .09       | .033   | 16.48 |
| CHC <sup>1a</sup> | ohne GKT     | 5,054    | 16       | -.03   | .059        | 56.67  | .05    | -.03      | .047   | 58.91 |
|                   | <i>PSK</i>   |          |          |        |             |        |        |           |        |       |
| CHC <sup>1</sup>  | NEO          | 4,831    | 19       | .01    | .089        | 38.04  | -.11   | .00       | .081   | 37.53 |
| CHC <sup>1</sup>  | NEO-PI-(R)   | 2,145    | 10       | .00    | .073        | 45,17  | -.10   | .00       | .076   | 44.98 |
| CHC <sup>1</sup>  | NEO-FFI      | 2,686    | 9        | .02    | .107        | 33.19  | -.11   | .00       | .085   | 31.70 |

*Anmerkungen.* *N* = Stichprobengröße; *k* = Anzahl inkludierter unabhängiger Datensätze;  $\rho$  = um den Stichprobenfehler, die Unreliabilität im Prädiktor und Kriterium, die Varianzunterschätzung und -überschätzung korrigierte Effektgröße;  $SD_{\rho}$  = Standardabweichung der artefaktkorrigierten Effektgröße; % Var. = erklärte Varianz; 90% CV = untere Grenze des Glaubwürdigkeitsintervall, aus der ersichtlich wird, ob die Null im Glaubwürdigkeitsintervall liegt;  $\bar{r}$  = ausschließlich um den Stichprobenfehler korrigierte Effektgröße („Bare-Bones“-Analysen); <sup>1</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde; <sup>1a</sup> = Datensätze, bei denen eine Zuordnung des Fähigkeitskonstruktes Gc durch den Verfasser der vorliegenden Arbeit im Nachhinein gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) vorgenommen wurde und die Studie von Reeve et al. (2006, *n* = 71,887) exkludiert ist; kogLT = kognitiver Leistungstest; PSK = Persönlichkeitsinventar; GKT = General Knowledge Test; die Studien inkludiert, die den GKT zur Erfassung der kristallisierten Intelligenzausprägung verwendet haben; ohne GKT = die Studien inkludiert, die zur Erfassung der kristallisierten Intelligenz andere kogLT als den GKT verwendet haben; NEO = die Studien inkludiert, in denen der NEO-PI-(R) und der NEO-FFI berücksichtigt wurden; NEO-PI-(R) = die Studien inkludiert, in denen der NEO-PI-R und der NEO-PI verwendet wurde.

Mit keinem Moderator konnte mehr als 56.67% der Varianz erklärt werden. Offenbar existieren weitere Moderatoren, die den Zusammenhang zwischen Gc und G erklären, jedoch im Rahmen der vorliegenden Arbeit nicht identifiziert werden konnten.

## *Kapitel 5 – Ergebnisse*

### *Zusammenfassung zum Zusammenhang zwischen der Kristallisierten Intelligenz und Gewissenhaftigkeit:*

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien vorgenommen wurde, beträgt die artefaktkorrigierte Effektgröße  $\rho = .13$ . Dabei kann von einem positiven Effekt in der Population ausgegangen werden, da das Glaubwürdigkeitsintervall die Null nicht mit einschließt. Nach der Exklusion der Studie von Reeve et al. (2006;  $n = 71,887$ ) beträgt die artefaktkorrigierte Effektgröße  $\rho = -.07$ . In dem Fall kann ein Nulleffekt in der Population nicht ausgeschlossen werden, da das Glaubwürdigkeitsintervall die Null mit einschließt. Unabhängig von der Exklusion bzw. Inklusion der Studie von Reeve et al. (2006) beträgt die erklärte Varianz weniger als 75%.

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) erfolgte, beträgt die artefaktkorrigierte Effektgröße  $\rho = .12$ . Hierbei kann von einem positiven Effekt in der Population ausgegangen werden, da das Glaubwürdigkeitsintervall die Null nicht mit einschließt. Nach der Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ) kann hingegen ein Nulleffekt in der Population bei einer artefaktkorrigierten Effektgröße von  $\rho = .00$  nicht ausgeschlossen werden. Unabhängig von der Exklusion bzw. Inklusion der Studie von Reeve et al. (2006) beträgt die erklärte Varianz in dieser Analyse-Variante ebenfalls weniger als 75%.

Hinsichtlich eines Publikationsbias kann in der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien vorgenommen wurde, von einer asymmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße von  $\bar{r} = .03$  mit  $R_0 = 1$  ausgegangen werden. Ein Publikationsbias sollte demnach nicht ausgeschlossen werden. Nach der Exklusion einer Effektgröße mit positiven Vorzeichen und dem höchsten Betrag sollte ein Publikationsbias mit  $R_0 < 1$  auszuschließen sein. Aufgrund der Höhe des zweiten Schätzer mit  $L_0 < 1$  kann ein Publikationsbias ausgeschlossen werden. Nach der durchgeführten Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens kann ein Publikationsbias ebenfalls ausgeschlossen werden.

In der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, kann auf Grund der Höhe der Schätzer  $R_0$  und  $L_0$  ein Publikationsbias nicht ausgeschlossen werden. Nach der Exklusion von einer positiven ( $R_0$ ) bzw. fünf negativen Effektgrößen ( $L_0$ ) kann gemäß der „trim & fill“-Methode ein Publikationsbias

## *Kapitel 5 – Ergebnisse*

ausgeschlossen werden. Nach der durchgeführten Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens kann ein Publikationsbias sowohl für den negativen als auch positiven Wertebereich nicht ausgeschlossen werden. Dabei sind jeweils zu viele signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert als man bei einer zufälligen Verteilung signifikanter ( $p \leq .05$ , zweiseitig) und nichtsignifikanter Effektgrößen ( $p > .05$ , zweiseitig) erwarten würde.

Mit keinem Moderator konnten mehr als 56.67% der Varianz erklärt werden. Offenbar existieren, neben dem kogLt und Persönlichkeitsinventar, weitere Moderatoren, die den Zusammenhang zwischen Gc und G erklären.

## **6 Zusammenfassung und Diskussion**

In der vorliegenden Arbeit wurden metaanalytische Untersuchungen zum Zusammenhang von den Fähigkeitskonstrukten Allgemeine Intelligenz (g), Fluide Intelligenz (Gf) und Kristallisierte Intelligenz (Gc) mit den Persönlichkeitsdimensionen Neurotizismus (Ne), Extraversion (E), Offenheit für Erfahrung (O), Verträglichkeit (V) und Gewissenhaftigkeit (G) durchgeführt. Dabei wurde von der Einschätzung ausgegangen, dass die Autoren der publizierten Studien eine z.T. liberale bzw. ungenaue Zuordnung der jeweiligen Intelligenztests bzw. kognitiven Leistungsaufgaben zu den Fähigkeitskonstrukten g, Gf oder Gc vorgenommen haben. Tatsächlich konnte beobachtet werden, dass die Zuordnung der kognitiven Leistungstests bzw. Intelligenztests zu den Fähigkeitskonstrukten in verschiedenen Studien nicht immer einheitlich war. Um diesbezügliche Ungenauigkeiten zu berücksichtigen und zu untersuchen, wurden zwei Analysen durchgeführt. In der ersten metaanalytischen Untersuchung wurde die Klassifizierung der Fähigkeitskonstrukte, wie sie von den Autoren der publizierten Studien erfolgte, beibehalten. Dadurch ist die Variabilität in der Zuordnung von Aufgaben zu Fähigkeitskonstrukten in der Variabilität der Effektstärken enthalten. Damit wird ein sozialkonstruktivistisches Wissenschaftsverständnis in der Metaanalyse abgebildet. Dennoch erscheint es auch angemessen, Befunde auf Basis einer einheitlichen Zuordnung zu generieren. Daher erfolgte in einer anschließenden, zweiten Analyse eine nachträgliche Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 1997, 2005, 2009) nach der Zuordnungsvorlage von Süß und Beauducel (2011) durch den Autor der vorliegenden Arbeit. Ein Vergleich der beiden Analysen erlaubt eine Einschätzung der Auswirkungen unterschiedlicher Zuordnungen der Fähigkeitskonstrukte auf die metaanalytischen Kennwerte. Darüber hinaus war es ein Anliegen der vorliegenden Arbeit, verschiedene Aspekte des Publikationsbias zu untersuchen. Dazu wurden zwei verschiedene Verfahren berücksichtigt. Mit der bereits publizierten und etablierten „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) wurde die symmetrische Verteilung der publizierten Effektgrößen untersucht. Waren die publizierten Effektgrößen nicht symmetrisch um die mittlere Effektgröße verteilt, sollte nach Duval und Tweedie (2000a, 2000b) und Duval (2005) ein Publikationsbias nicht ausgeschlossen werden. Dennoch kann es unter Umständen vorkommen, dass trotz symmetrisch verteilter Effektgrößen ein Publikationsbias nicht auszuschließen ist (siehe bspw. Abbildung 10, S. 51). Deshalb wurde eine zweite, im Rahmen der vorliegenden Arbeit entwickelte Methode angewendet, die die Möglichkeit bietet, ein Publikationsbias unabhängig von einer symmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße zu identifizieren. Unter Berücksichtigung verschiedener Parameter (siehe dazu Kapitel 4.6.7) wurde eine zufällige Verteilung signifikanter und nichtsignifikanter



Effektgrößen um die Populationseffektgröße – ohne Publikationsbias – simuliert. Anschließend wurde die simulierte Verteilung mit der Verteilung der tatsächlich publizierten signifikanten und nichtsignifikanten Effektgrößen – mit möglichen Publikationsbias – verglichen. Bestand eine Differenz zwischen dem simulierten Referenzwert und dem empirischen Wert, wurde ein Publikationsbias angenommen und die Anzahl der zusätzlichen Studien berechnet, die nötig wären, um ein Publikationsbias auszuschließen.

Insgesamt wurde(n) in der vorliegenden Arbeit

- metaanalytisch der Zusammenhang der Fähigkeitskonstrukte  $g$ ,  $G_f$  und  $G_c$  mit den Persönlichkeitsdimensionen des FFM untersucht,
- das Konstruktverständnis der Autoren publizierter Studien mit der CHC-Theorie (McCrew, 2009) hinsichtlich  $g$ ,  $G_f$  und  $G_c$  verglichen und
- verschiedene Aspekte des Publikationsbias untersucht.

Entsprechend diesem Ziel wurde in der vorliegenden Arbeit zunächst die Relevanz der Fähigkeitskonstrukte  $g$ ,  $G_f$  und  $G_c$  sowie der Persönlichkeitsdimensionen des FFM dargestellt und diskutiert (Kapitel 2.1; 2.2). Anschließend wurden sowohl unterschiedliche Perspektiven und Modelle als auch bisherige metaanalytische Ergebnisse zum Zusammenhang zwischen Intelligenz und Persönlichkeit erläutert (Kapitel 2.3.1 – 2.3.3). Daraus wurden die Begründungen zur Relevanz einer weiteren Metaanalyse zum Zusammenhang zwischen Intelligenz und Persönlichkeit abgeleitet (Kapitel 2.3.4). Darüber hinaus wurden verschiedene Aspekte des Publikationsbias beschrieben (Kapitel 2.3.5). Die Ziele der vorliegenden Arbeit sind im Kapitel 3 erläutert. In Kapitel 4 wurden die methodischen Überlegungen zur Durchführung metaanalytischer Untersuchungen sowie der Einsatz der Verfahren zur Identifikation eines Publikationsbias erläutert. Im Kapitel 5 wurden die metaanalytischen Ergebnisse zum Zusammenhang der Fähigkeitskonstrukte  $g$ ,  $G_f$  und  $G_c$  mit den Persönlichkeitsdimensionen des FFM, einschließlich der Resultate zum Publikationsbias, berichtet. Im vorliegenden Kapitel 6 wird versucht, die Ergebnisse der Arbeit in die bestehende Literatur einzuordnen. Anschließend werden die wesentlichen Einschränkungen und Grenzen der vorliegenden Arbeit und der weitere Forschungsbedarf diskutiert.

### **6.1 Ergebnisse der theoretischen und methodischen Überlegungen**

Um eine mögliche Variation der Populationsparameter in den verschiedenen Studien zuzulassen, wurde ein RE-Modell zur Bestimmung der metaanalytischen Kennwerte verwendet. Das RE-Modell ist aufgrund der Analyse, die die Variabilität der Zuordnung kognitiver Leistungstests zu  $g$ ,

Gf und Gc zulässt, unabdingbar. Von einer einheitlichen Population ist bei unterschiedlichen Zuordnungen der Aufgaben zu den Konstrukten nicht auszugehen. Durch die Verwendung des RE-Modells in der vorliegenden Arbeit ist eine Überschätzung der artefaktkorrigierten Effektgrößen mit hoher Wahrscheinlichkeit auszuschließen (vgl. Kapitel 4.3; Hunter & Schmidt, 2004).

Da eine Stabilität der Persönlichkeit im Alter unter 15 Jahren kontrovers diskutiert wurde (bspw. Eysenck & Cookson, 1969; Roberts & DelVecchio, 2000), wurde das Mindestalter der Probanden der inkludierten Studien in der vorliegenden Arbeit im Vergleich zu Ackerman und Heggestad (1997) sowie Wolf und Ackerman (2005) um fünf Jahre auf  $M_{\text{Alter}} - SD_{\text{Alter}} \geq 15$  Jahre erhöht. Unabhängig davon konnten, insbesondere bei der Untersuchung des Zusammenhangs von g, Gf und Gc mit den Persönlichkeitsdimensionen O, V und G, deutlich mehr Studien als bei Ackerman und Heggestad (1997) inkludiert werden. Aufgrund einer ausreichend großen Anzahl von inkludierten Studien können für alle in der vorliegenden Arbeit untersuchten Zusammenhänge robuste Schätzungen der Populationskennwerte angenommen werden.

Neben der Realisierung von zwei metaanalytischen Untersuchungen aufgrund unterschiedlicher Zuordnungen der Fähigkeitskonstrukte g, Gf und Gc – gemäß Autoren publizierter Studien vs. CHC-Theorie – wurde der Einfluss von Studien mit sehr großen Stichprobengrößen kontrolliert. Dem Vorschlag von Hunter und Schmidt (2004) folgend, wurden jeweils zwei Varianten gerechnet. Zuerst blieben die Studien von Reeve et al. (2006;  $n = 71,887$ ) oder De Fruyt et al. (2006;  $n = 20,175$ ) inkludiert, anschließend wurden diese Studien exkludiert. Der Einfluss von Studien mit großen Stichprobengrößen auf die metaanalytischen Kennwerte konnte so gezeigt werden. Allerdings gibt es keine standardisierte Vorgehensweise, wie zu verfahren ist, wenn sich das Ergebnis durch die Exklusion einer einzelnen Studie substantiell ändert (vgl. Hunter & Schmidt, 2004).

Rothstein und Bushman (2012) schlagen vor, möglichst verschiedene Methoden zur Identifikation eines Publikationsbias zu nutzen. Dementsprechend wurden in der vorliegenden Arbeit zwei Methoden verwendet, um unterschiedliche Aspekte des Publikationsbias zu untersuchen. Bei einer symmetrischen Verteilung publizierter Effektgrößen um die mittlere Effektgröße kann nach der „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) ein Publikationsbias ausgeschlossen werden. Allerdings kann eine symmetrische Verteilung publizierter Effektgrößen unzureichend sein, um ein Publikationsbias zu identifizieren. Deshalb wurde zusätzlich unter Berücksichtigung der Signifikanzschwelle und mittels Simulation eine im Rahmen der

vorliegenden Arbeit entwickelte Methode angewendet, um die Wahrscheinlichkeit zu erhöhen, ein Publikationsbias ungeachtet der symmetrischen Verteilung publizierter Effektgrößen zu identifizieren.

## **6.2 Ergebnisse und Diskussion zum Zusammenhang zwischen Intelligenz und Persönlichkeit**

### *Der Zusammenhang zwischen Intelligenz und Neurotizismus:*

Anhand der Literatur kann eine negative Beziehung zwischen Neurotizismus und Intelligenz angenommen werden (bspw. Stumm et al., 2011; Zeidner & Matthews, 2000). Diese negative Beziehung soll darauf zurückzuführen sein, dass Individuen mit einer höheren Neurotizismusausprägung während einer Testsituation nervöser und ängstlicher sind, weshalb ihre Performanz in kognitiven Leistungstests geringer ausfällt (Chamorro-Premuzic & Furnham, 2004; Moutafi et al., 2006; Zeidner & Matthews, 2000). Moutafi et al. (2006) konnten darüber hinaus zeigen, dass nachdem die Testangst herauspartialisiert wurde, keine negative Beziehung mehr zwischen Neurotizismus und Intelligenz zu beobachten war. Diese Annahmen und Ergebnisse werden durch Untersuchungen gestützt, in denen gezeigt wurde, dass Individuen mit einer höheren Neurotizismusausprägung insgesamt sensibler auf Stress reagieren (Bolger & Schilling, 1991; Gross, Sutton & Ketelaar, 1998; Robinson & Tamir, 2005). Gleichzeitig soll eine höhere Neurotizismusausprägung mit erhöhtem „cognitive noise“ einhergehen (Robinson & Tamir, 2005). Dabei kann vermutet werden, dass in einer Testsituation mehr kognitive Ressourcen gebunden sind, um das Stresslevel zu kontrollieren, was sich leistungsmindernd auf das Testresultat auswirkt. Bisherige metaanalytische Ergebnisse unterstützen diese Annahmen, da negative Populationszusammenhänge zwischen  $g$  und  $Ne$  berichtet wurden (Ackerman & Heggestad, 1997; Judge et al., 2007; Poropat, 2009).

Zwischen  $g$  und  $Ne$  kann in der hier durchgeführten Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien berücksichtigt wurde, von einem negativen Populationszusammenhang ausgegangen werden. Nach der anschließenden Exklusion einer einzelnen Studie mit einer Stichprobengröße von  $n = 71,887$  (Reeve et al., 2006) kann ein Nulleffekt in der Population nicht mehr ausgeschlossen werden. Hierbei ist es beachtenswert, dass sich bereits durch die Exklusion einer einzelnen Studie bei einer inkludierten Stichprobenanzahl von immerhin  $k = 69$  der Befund hinsichtlich eines Nulleffektes substantiell verändert. Vermutlich sind diese Befunde von einem z.T. liberaleren Konstruktverständnis von  $g$  durch die Autoren der publizierten Studien beeinflusst. Denn in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte einheitlich gemäß der CHC-Theorie (McGrew, 2009) im Nachhinein

erfolgte, kann ein negativer Populationszusammenhang zwischen  $g$  und  $N_e$ , unabhängig von der Exklusion bzw. Inklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ), angenommen werden. Darüber hinaus ist sowohl nach der „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) als auch nach der Simulationsmethode ein Publikationsbias in der Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien erfolgte, anzunehmen. Offenbar sind mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) mit negativem Vorzeichen im Vergleich zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde. Dementsprechend ist davon auszugehen, dass Studien mit nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) im Vergleich zu signifikanten Effektgrößen ( $p \leq .05$ , zweiseitig) häufiger nicht zur Publikation gelangt sind. Hinsichtlich der Moderatoranalysen kann sowohl für Eysenck-Inventare als auch NEO-FFI-Inventare von einem negativen Populationseffekt zwischen  $g$  und  $N_e$  ausgegangen werden. Hingegen kann für NEO-PI-R-Inventare ein Nulleffekt zwischen  $g$  und  $N_e$  nicht ausgeschlossen werden.

Wurde in Ackerman und Heggstad (1997) eine negative Populationseffektgröße zum Zusammenhang zwischen  $G_f$  und  $N_e$  berichtet, kann in der vorliegenden Arbeit, unabhängig von der Zuordnung der Fähigkeitskonstrukte ein Populationseffekt von Null nicht ausgeschlossen werden. Dieses Ergebnis änderte sich auch nicht durch die Exklusion von De Fruyt et al. (2006) mit einer Stichprobengröße von  $n = 20,175$ . Die Veränderung von einem negativen Populationseffekt (Ackerman & Heggstad, 1997) zu einem Nulleffekt in der vorliegenden Arbeit kann vermutlich auf den zunehmenden Einsatz von NEO-Inventaren, der mit einer Verdrängung von Eysenck-Inventaren einherzugehen scheint, zurückgeführt werden. In der vorliegenden Arbeit konnten entsprechend den Inklusionskriterien 39 Studien, die ein NEO-Inventar und 10 Studien, die ein Eysenck-Inventar zur Erfassung der Neurotizismusaussprägung eingesetzt hatten, für die metaanalytische Auswertung berücksichtigt werden. Für die Eysenck-Inventare beträgt die artefaktkorrigierte Effektgröße in der vorliegenden Arbeit  $\rho = -.08$ . Dabei ist die Höhe der Populationseffektgröße des Zusammenhangs zwischen  $G_f$  und  $N_e$  identisch mit der aus Ackerman und Heggstad (1997). Hingegen kann für die NEO-Inventare ein Nulleffekt aufgrund der Ergebnisse der vorliegenden Arbeit nicht ausgeschlossen werden. Offenbar gibt es konzeptionelle Unterschiede zwischen Eysenck- und FFM-Inventaren, die in Bezug auf Neurotizismus relevant für die Zusammenhänge mit  $G_f$  sind. Möglicherweise sind in den Neurotizismus-Items aus den Eysenck-Inventaren mehr klinische Aspekte enthalten (bspw. „Haben Sie es mit den Nerven?“ oder „Leiden Sie an Schlaflosigkeit?“) als in den Neurotizismus-Items aus dem NEO-FFI (bspw.

„Zu häufig bin ich entmutigt und will aufgeben, wenn etwas schiefgeht“ oder „Ich bin nicht leicht beunruhigt“), die sich unterschiedlich auf die kognitive Leistungsfähigkeit, insbesondere Gf, auswirken.

Ein Publikationsbias kann nach der „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien berücksichtigt wurde, ausgeschlossen werden. Hingegen sollte in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, ein Publikationsbias nicht ausgeschlossen werden, da sich die publizierten Effektgrößen nicht symmetrisch um die mittlere Effektgröße verteilen. Mittels durchgeführter Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens kann unabhängig der Analyse-Variante – Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien vs. CHC-Theorie – ein Publikationsbias zum Zusammenhang zwischen Gf und Ne nicht ausgeschlossen werden. Offenbar wurden mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) mit negativen Vorzeichen im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde. Dementsprechend ist davon auszugehen, dass nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) z.T. nicht zur Publikation gelangt sind.

Analog zu Ackerman und Heggstad (1997) kann in der vorliegenden Arbeit, unabhängig von der Zuordnung der Fähigkeitskonstrukte und der Exklusion der Studie von Reeve et al. (2006,  $n = 71,887$ ), von einem negativen Populationszusammenhang zwischen Gc und Ne ausgegangen werden. Darüber hinaus kann ein Publikationsbias in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien vorgenommen wurde, ausgeschlossen werden. Allerdings beträgt die inkludierte Studienzahl lediglich  $k = 9$ , was als Einschränkung bei der Interpretation eines Publikationsbias angesehen werden sollte. Insofern ist eine inkludierte Studienzahl von  $k = 29$  in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, aussagekräftiger, um ein Publikationsbias einzuschätzen. Zumal in dieser Analyse-Variante sowohl die Simulationsmethode als auch die „trim & fill“-Methode ein Publikationsbias anzeigt. Dabei kann davon ausgegangen werden, dass mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) mit negativen Vorzeichen im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert sind, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde.

## *Kapitel 6 – Zusammenfassung und Diskussion*

Offenbar sind Studien aufgrund nichtsignifikanter Effektgrößen ( $p > .05$ , zweiseitig) z.T. nicht zur Publikation gelangt.

### *Der Zusammenhang zwischen Intelligenz und Extraversion:*

In der Literatur werden verschiedene Ergebnisse und Annahmen zum Zusammenhang zwischen Intelligenz und Extraversion berichtet (bspw. Ackerman, 2000; Ackerman & Heggestad, 1997; Moutafi et al., 2003; Rawlings & Carnie, 1989; Rawlings & Skok, 1993; Roberts, 2002; Robinson, 1985; Rolfhus & Ackerman, 1999; Saklofske & Kostura, 1990; Wolf & Ackerman, 2005; Zeidner & Matthews, 2000). Die Höhe des Zusammenhangs zwischen Intelligenz und Extraversion kann mit der kortikalen Erregung und dem Yerkes-Dodson-Gesetz begründet werden (Eysenck, 1967; Yerkes & Dodson, 1908). Das Yerkes-Dodson-Gesetz besagt, dass eine mittlere Erregung („intermediate level of arousal“) zu einer optimalen Performanz führt (Yerkes & Dodson, 1908). Weiterhin berichtet Eysenck (1967, 1994), dass Individuen mit einer höheren Extraversionausprägung eine dispositionale niedrigere kortikale Erregung aufweisen sollen als Individuen mit einer geringeren Extraversionausprägung (Eysenck, 1967; Eysenck, 1994). Hinsichtlich der zuvor beschriebenen Annahmen sollen kognitive Leistungstests mit zeitlicher Begrenzung Vorteile für Individuen mit einer höheren Extraversionausprägung bieten (Chamorro-Premuzic & Furnham, 2004; Rawlings & Carnie, 1989; Zeidner & Matthews, 2000). Diese Autoren gehen davon aus, dass durch den Zeitdruck die niedrige kortikale Erregung zunimmt, was zu einer Erhöhung der Aufmerksamkeit, Wachsamkeit und Reaktionsbereitschaft führen kann und sich positiv auf die Testresultate auswirkt. Hingegen kann sich bei Individuen mit niedrigerer Extraversionausprägung die zeitliche Begrenzung in kognitiver Leistungsminderung auf das Testergebnis auswirken, da die kortikale Erregung zu hoch ist (Rawlings & Carnie, 1989). Eine weitere Annahme ist, dass sich die kognitive Leistungsfähigkeit in Abhängigkeit der Extraversionausprägung auf unterschiedliche Aufgabentypen auswirken kann (bspw. Robinson, 1985). Dabei wird davon ausgegangen, dass Individuen mit einer höheren Extraversionausprägung besser in Aufgabentypen abschneiden, in denen motorische Fähigkeiten zählen (Robinson, 1985). Allerdings konnten sowohl die Befunde aus Rawlings und Carnie (1989) als auch Robinson (1985) nicht erfolgreich repliziert werden (bspw. Rawlings & Skok, 1993; Saklofske & Kostura, 1990). Hinsichtlich metaanalytischer Untersuchungen zum Zusammenhang zwischen Intelligenz und Extraversion konnten Wolf und Ackerman (2005) eine Abnahme von einem positiven zu einem negativen Populationszusammenhang während der letzten Jahre feststellen. Als eine der wesentlichen Gründe für diese Abnahme des Populationszusammenhangs wird der zunehmende Einsatz von NEO-Inventaren, der mit einer Verdrängung von Eysenck-

Inventaren einherzugehen scheint, genannt (Wolf & Ackerman, 2005). Möglicherweise erfassen Extraversions-Items aus den Eysenck-Inventaren Anteile, die für die kognitive Leistungsfähigkeit im Vergleich zu den Extraversions-Items aus den NEO-Inventaren relevant sind.

Anhand der Ergebnisse aus der vorliegenden Arbeit kann unabhängig von der Zuordnung der Fähigkeitskonstrukte – gemäß Autoren publizierter Studien vs. CHC-Theorie – und unabhängig von der Exklusion oder Inklusion der Studie von Reeve et al. (2006) mit  $n = 71,887$  ein Effekt von Null in der Population zwischen  $g$  und  $E$  nicht ausgeschlossen werden. Damit unterscheidet sich das Ergebnis der vorliegenden Arbeit von den Ergebnissen in Ackerman und Heggestad (1997) sowie Wolf und Ackerman (2005), wo jeweils ein positiver Effekt zwischen  $g$  und  $E$  berichtet wurde. Die metaanalytischen Untersuchungen von Judge et al. (2007) und Poropat (2009) erlauben keine Interpretation der artefaktkorrigierten Effektgröße zum Zusammenhang zwischen  $g$  und  $E$ , da weder das Glaubwürdigkeitsintervall noch die Standardabweichung der artefaktkorrigierten Effektgröße berichtet wurden. Allerdings ist bei artefaktkorrigierten Effektgrößen von  $\rho = .02$  (Judge et al., 2007) und  $\rho = -.01$  (Poropat, 2009) die Wahrscheinlichkeit hoch, dass ein Nulleffekt in der Population vorliegt.

Die Befunde aus der vorliegenden Arbeit zum Zusammenhang zwischen  $g$  und  $E$  unterstützen die Einschätzung von Wolf und Ackerman (2005), dass ein zunehmender Einsatz von NEO-Inventaren bei einer gleichzeitigen Verdrängung von Eysenck-Inventaren über die Zeit einherzugehen scheint. Das erste NEO-Inventar wurde im Jahr 1985 publiziert. Dies war ein Grund, das Jahr 1985 als untere zeitliche Inklusionsgrenze potenzieller Studien im Rahmen der vorliegenden Arbeit zu berücksichtigen (siehe Kapitel 4.2). In der vorliegenden Arbeit wurden unter Berücksichtigung der Inklusionskriterien 47 Studien, die Extraversion mit einem NEO-Inventar und 11 Studien, die Extraversion mit einem Eysenck-Inventar erfasst haben, inkludiert. Hier zeigte sich eine deutliche Dominanz von NEO-Inventaren gegenüber Eysenck-Inventaren. Hingegen konnten die Befunde eines negativen Zusammenhangs aus Wolf und Ackerman (2005) für die Moderatoren NEO-FFI und NEO-PI-R nicht bestätigt werden. Aufgrund der Ergebnisse der vorliegenden Arbeit kann ein Nulleffekt in der Population nicht ausgeschlossen werden. Hinsichtlich eines Publikationsbias zum Zusammenhang zwischen  $g$  und  $E$  kann von einer asymmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße ausgegangen werden. Ein Publikationsbias sollte demnach nicht ausgeschlossen werden. Die Ergebnisse der Simulationsmethode zeigten ebenfalls ein Publikationsbias an. Offenbar sind mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p$

> .05, zweiseitig) publiziert, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde. Dementsprechend ist davon auszugehen, dass nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) häufiger nicht zur Publikation gelangt sind als signifikante Effektgrößen ( $p \leq .05$ , zweiseitig).

Wie bereits zum Zusammenhang zwischen  $g$  und  $E$  kann auch zwischen  $G_f$  und  $E$ , unabhängig von der Zuordnung der Fähigkeitskonstrukte und unabhängig von der Exklusion bzw. Inklusion der Studie von De Fruyt et al. (2006) mit  $n = 20,175$ , ein Nulleffekt in der Population nicht ausgeschlossen werden. Damit unterscheidet sich das Ergebnis der vorliegenden Arbeit von den Ergebnissen aus Ackerman und Heggestad (1997) sowie Wolf und Ackerman (2005), die einen positiven Populationszusammenhang berichten. In keiner, der in der vorliegenden Arbeit gerechneten Modertoranalysen kann von einem positiven Populationszusammenhang zwischen  $G_f$  und  $E$  ausgegangen werden.

Nach der „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) sollte in beiden Analyse-Varianten – Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien vs. CHC-Theorie – ein Publikationsbias nicht ausgeschlossen werden, da sich die publizierten Effektgrößen nicht symmetrisch um die mittlere Effektgröße verteilen. Ebenfalls sollte nach der Simulationsmethode ein Publikationsbias in beiden Analyse-Varianten nicht ausgeschlossen werden. Mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) waren im Vergleich zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde. Entsprechend ist davon auszugehen, dass Studien aufgrund nichtsignifikanter Effektgrößen z.T. nicht zur Publikation gelangt sind.

Zum Zusammenhang zwischen  $G_c$  und  $E$  kann in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien erfolgte, ein positiver Effekt in der Population angenommen werden. Nach der Exklusion der Studie von Reeve et al. (2006) mit einer Stichprobengröße von  $n = 71,887$  konnte ein Nulleffekt in der Population nicht mehr ausgeschlossen werden. In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, kann unabhängig von der Exklusion bzw. Inklusion der Studie von Reeve et al. (2006; mit  $n = 71,887$ ) ein Nulleffekt in der Population nicht ausgeschlossen werden. Der Populationszusammenhang zwischen  $G_c$  und  $E$  beträgt in Ackerman und Heggestad (1997)  $\rho = .13$ . Eine leichte Abnahme ist in Wolf und



## *Kapitel 6 – Zusammenfassung und Diskussion*

Ackerman (2005) zum Zusammenhang zwischen Gc und E zu beobachten ( $\rho = .08$ ). In der vorliegenden Arbeit kann ein Nulleffekt in der Population nicht mehr ausgeschlossen werden. Es kann vermutet werden, dass der positive Zusammenhang zwischen Gc und E über die Zeit abnimmt (vgl. Wolf & Ackerman, 2005). Möglicherweise ist dies auf den häufigeren Einsatz von NEO-Inventaren zurückzuführen: Wie bereits zum Zusammenhang zwischen Gf und E kann auch hier in den Studien, in denen der NEO-FFI zur Erfassung von E eingesetzt wurde, von einem negativen Zusammenhang ausgegangen werden.

Ein Publikationsbias kann nach der „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) ausgeschlossen werden, da sich die publizierten Effektgrößen symmetrisch um die mittlere Effektgröße verteilt haben. Hingegen kann nach der Simulation gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens ein Publikationsbias nicht ausgeschlossen werden. Mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) als nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) sind publiziert, als man bei einer zufälligen Verteilung signifikanter ( $p \leq .05$ , zweiseitig) und nichtsignifikanter Effektgrößen ( $p > .05$ , zweiseitig) um die Populationseffektgröße erwarten würde. Dementsprechend ist davon auszugehen, dass Studien aufgrund nichtsignifikanter Effektgrößen vermehrt nicht zur Publikation gelangt sind.

Zusammenfassend kann zum Zusammenhang von den drei Intelligenzkonstrukten g, Gf, Gc mit der Persönlichkeitsdimension Extraversion festgehalten werden: Neuere Studien, die nach den Arbeiten von Ackerman und Heggestad (1997) sowie Wolf und Ackerman (2005) publiziert wurden, berichten vermehrt Effektgrößen nahe Null. Darüber hinaus konnten in den Studien, in denen der NEO-FFI eingesetzt wurde, häufiger negative Zusammenhangsmaße zwischen den Intelligenzkonstrukten g, Gf, Gc und E beobachtet werden. Offenbar existieren konzeptionelle Unterschiede zwischen dem NEO-FFI und Nicht-NEO-FFI-Inventaren für Extraversion, die bezüglich der untersuchten Intelligenzkonstrukte von Relevanz zu sein scheinen.

### *Der Zusammenhang zwischen Intelligenz und Offenheit für Erfahrungen:*

Offenheit für Erfahrung ist die Persönlichkeitsdimension des FFM, die am höchsten mit der Intelligenz positiv korreliert (e.g., Ackerman & Heggestad, 1997; Furnham et al., 2005; Judge et al., 2007; Poropat, 2009; Zeidner & Matthews, 2000). In der Literatur existiert eine Vielzahl von Annahmen bzw. Modellen, die den Zusammenhang zwischen Offenheit für Erfahrungen und Intelligenz beschreiben (bspw. Ackerman, 1996; Ackerman & Goff, 1994; Beauducel et al., 2007; Brand, 1994; Cattell, 1987; Chamorro-Premuzic & Arteché, 2008; DeYoung et al., 2005; Raine,

Reynolds, Venables & Mednick, 2002; Zeidner & Matthews, 2000; Ziegler, Danay, Heene, Asendorpf & Bühner, 2012). Nach der „Environmental Enrichment Hypothesis“ (bspw. Raine et al., 2002) wird angenommen, dass Gf durch O über die Zeit beeinflusst wird. Demnach soll eine höhere Offenheitsausprägung zu mehr Lerngelegenheiten führen, was nicht nur zu einer Zunahme von Gc, sondern auch zu einer Zunahme von Gf führt. Nach der „Environmental Success Hypothesis“ (bspw. DeYoung et al., 2005) wird angenommen, dass Gf die Persönlichkeitsdimension Offenheit für Erfahrungen beeinflusst. Personen mit einer höheren Gf-Ausprägung sollen neue Herausforderungen leichter bzw. erfolgreicher bewältigen als Personen, die eine geringere Gf-Ausprägung haben. Diese damit verbundenen positiven Erfahrungen der Personen mit höheren Gf-Werten sollen sich in einer Zunahme des Interesses und der Neugier zeigen und zu einer Zunahme der Offenheit für Erfahrungen über die Zeit führen (Beauducel et al., 2007; Chamorro-Premuzic & Furnham, 2004; Holland, Dollinger, Holland & MacDonald, 1995; Ziegler et al., 2012). Der positive Populationszusammenhang in der vorliegenden Arbeit zwischen Gf und O und im Gegensatz zu Ackerman und Heggestad (1997) unterstützt sowohl die „Environmental Enrichment Hypothesis“ (bspw. Raine et al., 2002) als auch die „Environmental Success Hypothesis“ (bspw. DeYoung et al., 2005), ohne die Richtung der Beeinflussung zwischen Gf und O aufgrund korrelationsbasierter metaanalytischer Ergebnisse zu klären.

Durch den positiven Populationszusammenhang zwischen Gc und O und analog zu Ackerman und Heggestad (1997) kann die Annahme, dass eine höhere Offenheits-Ausprägung zu einer stärkeren Nutzung von Lerngelegenheiten und deshalb auch zu einer Zunahme von Gc im Sinne der „Environmental Enrichment Hypothesis“ führt (bspw. Ackerman, 1996; Ashton et al., 2000; Ziegler et al., 2012), unterstützt werden.

Hinsichtlich des Publikationsbias kann in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien vorgenommen wurde, nach der „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) ein Publikationsbias ausgeschlossen werden. Hingegen kann ein Publikationsbias in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, nicht ausgeschlossen werden, da sich die publizierten Effektgrößen asymmetrisch um die mittlere Effektgröße verteilt haben.

Zum Zusammenhang zwischen Gf und O kann in beiden Analyse-Varianten – Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien vs. CHC-Theorie – ein

Populationszusammenhang von Null nicht ausgeschlossen werden. Dieser Effekt ist jedoch durch eine einzelne Studie mit einer Stichprobengröße von  $n = 20,175$  (De Fruyt et al., 2006) bedingt. Nach der Exklusion dieser Studie kann für beide Analyse-Varianten ein positiver Zusammenhang zwischen Gf und O angenommen werden. Dieses Ergebnis unterscheidet sich von Ackerman und Heggestad (1997), in der ein Nulleffekt in der Population nicht auszuschließen ist. Allerdings beträgt die inkludierte Studienzahl in Ackerman und Heggestad (1997) lediglich  $k = 2$  bei einer Stichprobengröße von  $N = 205$ , weshalb die Robustheit des Populationszusammenhangs als gering im Vergleich zu dem Populationszusammenhang der vorliegenden Arbeit bewerten werden kann ( $k = 50$ ;  $N = 19,324$ ).

Hinsichtlich des Publikationsbias konnte nach der „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) in beiden Analyse-Varianten ein Publikationsbias zum Zusammenhang zwischen Gf und O nicht ausgeschlossen werden. Mittels durchgeführter Simulationen gemäß des im Kapitel 4.6.7 beschriebenen Vorgehens konnte ebenfalls ein Publikationsbias nicht ausgeschlossen werden. Offenbar sind die publizierten Effektgrößen asymmetrisch um die mittlere Effektgröße verteilt. Außerdem sind mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Vergleich zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde. Entsprechend ist davon auszugehen, dass Studien aufgrund nichtsignifikanter Effektgrößen z.T. nicht zur Publikation gelangt sind.

Zum Zusammenhang zwischen Gc und O kann aufgrund der metaanalytischen Kennwerte und analog zu Ackerman und Heggestad (1997) von einem positiven Populationszusammenhang ausgegangen werden. Nach der „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) sollte ein Publikationsbias aufgrund einer asymmetrischen Verteilung publizierter Effektgrößen nicht ausgeschlossen werden. Mittels durchgeführter Simulationen sollte in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde, ebenfalls ein Publikationsbias nicht ausgeschlossen werden. Dabei waren mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Vergleich zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde. Allerdings sollte bei Beurteilung hinsichtlich eines Publikationsbias die inkludierte Studienanzahl von lediglich  $k = 6$  berücksichtigt werden. Um robustere Ergebnisse zum Publikationsbias in dieser Analyse-Variante zu erhalten, sind weitere Publikationen abzuwarten. In der Analyse-Variante, in der eine

## *Kapitel 6 – Zusammenfassung und Diskussion*

Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) mit  $k = 33$  vorgenommen wurde, sollte ein Publikationsbias ausgeschlossen werden, da die Verteilung publizierter signifikanter ( $p \leq .05$ , zweiseitig) und nichtsignifikanter Effektgrößen ( $p > .05$ , zweiseitig) mit der simulierten Verteilung signifikanter ( $p \leq .05$ , zweiseitig) und nichtsignifikanter Effektgrößen ( $p > .05$ , zweiseitig) – ohne Publikationsbias – übereinstimmt.

Mit keinem in der vorliegenden Arbeit untersuchten Moderator konnten mehr als 75% der Varianz zum Zusammenhang zwischen g, Gf, Gc und O erklärt werden. Daher kann davon ausgegangen werden, dass weitere Moderatoren existieren, die den Zusammenhang zwischen g, Gf, Gc und O erklären. Aufgrund der Vielzahl von Annahmen bzw. Theorien, wie sich Intelligenz und Offenheit für Erfahrungen über die Zeit beeinflussen (bspw. Ackerman, 1996; Ackerman & Goff, 1994; Brand, 1994; Cattell, 1987; Chamorro-Premuzic & Arteche, 2008; DeYoung et al., 2005; Raine et al., 2002; Zeidner & Matthews, 2000; Ziegler et al., 2012), kann vermutet werden, dass der Moderator „Lebensalter der Personen in Jahren“ einen substantiellen Einfluss auf den Zusammenhang zwischen Intelligenz und Offenheit für Erfahrungen hat (siehe dazu auch 6.3).

### *Der Zusammenhang zwischen Intelligenz und Verträglichkeit:*

Analog zu den Ergebnissen aus Ackerman und Heggestad (1997) sowie Judge et al. (2007) kann zum Zusammenhang zwischen g und V ein Nulleffekt in der Population nicht ausgeschlossen werden. In der durchgeführten Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, kann von einem positiven Effekt ausgegangen werden. Allerdings wurde dieser positive Effekt durch eine einzelne Studie von Reeve et al. (2006,  $n = 71,887$ ) bei einer inkludierten Studienzahl von immerhin  $k = 59$  moderiert. Nach der anschließenden Exklusion dieser Studie konnte ein Nulleffekt zum Zusammenhang zwischen g und V in der Population nicht ausgeschlossen werden.

Nach der „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) kann in beiden Analyse-Varianten – Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien vs. CHC-Theorie – ein Publikationsbias nicht ausgeschlossen werden, da sich die publizierten Effektgrößen asymmetrisch um die mittlere Effektgröße verteilen. Aufgrund der Ergebnisse der Simulationen sollte ebenfalls ein Publikationsbias in beiden Analyse-Varianten nicht ausgeschlossen werden. Es waren mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) im Verhältnis zu nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) publiziert, als man bei einer zufälligen Verteilung der Effektgrößen um die Populationseffektgröße erwarten würde. Dabei

kann ein Publikationsbias sowohl für den negativen als auch positiven Wertebereich nicht ausgeschlossen werden. Dementsprechend kann davon ausgegangen werden, dass – unabhängig vom Vorzeichen des Zusammenhangsmaßes – Studien aufgrund nichtsignifikanter Effektgrößen ( $p > .05$ , zweiseitig) z.T. nicht zur Publikation gelangt sind. Da es zum Zusammenhang zwischen  $g$  und  $V$  keine erwartete theoretische Richtung des Zusammenhangs gibt (vgl. Ackerman, 1996; Ackerman & Heggestad, 1997; Beauducel et al., 2007; Cattell, 1987; Chamorro-Premuzic & Arteche, 2008; Goff & Ackerman, 1992), scheint das Vorzeichen der Effektgrößen zur Publikation vernachlässigbar zu sein, jedoch nicht die Signifikanz (siehe auch 6.3).

Mit keinem Moderator konnten mehr als 75% der Varianz zum Zusammenhang zwischen  $g$  und  $V$  erklärt werden. Offenbar existieren weitere Subpopulationen bzw. Moderatoren, die den Zusammenhang substantiell erklären.

Zum Zusammenhang zwischen  $G_f$  und  $V$  kann in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien vorgenommen wurde, von einem kleinen positiven Populationszusammenhang von  $\rho = .02$  ausgegangen werden. In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) erfolgte, kann hingegen ein Nulleffekt in der Population nicht ausgeschlossen werden. Hinsichtlich eines Publikationsbias ist nach der „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) in beiden Analyse-Varianten ein Publikationsbias tendenziell nicht auszuschließen, da eine asymmetrische Verteilung der publizierten Effektgrößen um die mittlere Effektgröße angezeigt wurde. Mittels Simulation kann hingegen in beiden Analyse-Varianten ein Publikationsbias ausgeschlossen werden. Die Verteilung publizierter signifikanter ( $p \leq .05$ , zweiseitig) und nichtsignifikanter Effektgrößen ( $p > .05$ , zweiseitig) stimmt mit der Verteilung der simulierten Effektgrößen – ohne Publikationsbias – überein. In Abhängigkeit der verwendeten Methode erhält man zum Zusammenhang zwischen  $G_f$  und  $V$  unterschiedliche Ergebnisse hinsichtlich des Vorhandenseins eines Publikationsbias. Der Vorschlag von Rothstein und Bushman (2012), möglichst verschiedene Methoden zur Identifikation eines Publikationsbias zu verwenden, erscheint nicht zuletzt auch hier sinnvoll, um komplexe Verteilungen der publizierten Effektgrößen zu untersuchen und möglichst differenziert zu bewerten: Zum Zusammenhang zwischen  $G_f$  und  $V$  kann davon ausgegangen werden, dass sich die publizierten Effektgrößen asymmetrisch um die mittlere Effektgröße verteilen. Dennoch gibt es hinsichtlich der Verteilung signifikanter und nichtsignifikanter Effektgrößen keinen Hinweis auf ein Publikationsbias.

Zum Zusammenhang zwischen Gc und V kann in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien vorgenommen wurde, von einem positiven Zusammenhang in der Population ausgegangen werden. Dieser Befund wird allerdings durch eine einzelne Studie von Reeve et al. (2006) mit einer Stichprobe von  $n = 71,887$  moderiert. Durch die Exklusion der Studie von Reeve et al. (2006) kann von einem kleinen negativen Populationszusammenhang ( $\rho = -.02$ ) ausgegangen werden. In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, kann von einem positiven Zusammenhang in der Population ausgegangen werden. Allerdings wurde auch dieses Ergebnis durch die Studie von Reeve et al. (2006,  $n = 71,887$ ) moderiert. Durch die Exklusion dieser Studie konnte ein Nulleffekt in der Population nicht mehr ausgeschlossen werden. Aufgrund der inkludierten Studienzahl von  $k = 20$  sind die Ergebnisse der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, als robuster zu bewerten im Vergleich zu der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der publizierten Studien mit  $k = 7$  erfolgte.

Hinsichtlich eines Publikationsbias konnte nach der „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien vorgenommen wurde, ein Publikationsbias ausgeschlossen werden. Hingegen konnte in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, ein Publikationsbias, aufgrund der Höhe des Schätzers mit  $R_0 = 1$ , nicht ausgeschlossen werden. Gemäß der Simulation konnte konsistent in beiden Analyse-Varianten ein Publikationsbias ausgeschlossen werden.

Bezüglich der Moderatoranalysen zeigte sich, dass der Zusammenhang zwischen Gc und V in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien berücksichtigt wurde, vollständig durch die Studie von Reeve et al. (2006,  $n = 71,887$ ) moderiert wurde. Die erklärte Varianz erhöhte sich nach der Exklusion von Reeve et al. (2006) von 8.75% auf 100%. Dementsprechend wurden keine weiteren Moderatoranalysen zum Zusammenhang zwischen Gc und V in dieser Analyse-Variante durchgeführt. In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, erhöhte sich die erklärte Varianz durch die Exklusion der Studie von Reeve et al. (2006) von 12.81% auf immerhin 51.36%. In anschließenden Moderatoranalysen wurde in den Studien, in denen der GKT zur Erfassung von Gc und/oder der NEO-PI-R zur Erfassung von V eingesetzt

## *Kapitel 6 – Zusammenfassung und Diskussion*

wurde, 100% der Varianz mit einer artefaktkorrigierten Effektgröße von jeweils  $\rho = .04$  erklärt. Offenbar ist der GKT und der NEO-PI-R relevant für den Zusammenhang zwischen Gc und V. In den Studien, in denen nicht der GKT oder NEO-PI-R eingesetzt wurde, betrug die erklärte Varianz deutlich weniger als 75%. Es kann davon ausgegangen werden, dass weitere Moderatoren existieren, die den Zusammenhang zwischen Gc und V erklären.

### *Der Zusammenhang zwischen Intelligenz und Gewissenhaftigkeit:*

In einer Vielzahl von Studien wurde die Relevanz des Einflusses der Persönlichkeitsdimension Gewissenhaftigkeit auf die Performanz im akademischen und beruflichen Kontext gezeigt (bspw. Blickle, 1996; Chamorro-Premuzic & Arteché, 2008; Schmidt & Hunter, 1998). Allerdings ist die Höhe des Populationszusammenhangs zwischen der kognitiven Leistungsfähigkeit und Gewissenhaftigkeit als gering zu bewerten (vgl. Ackerman & Heggestad, 1997; Judge et al., 2007; Poropat, 2009). Nichtsdestotrotz gibt es Annahmen bzw. Theorien zum Zusammenhang zwischen der kognitiven Leistungsfähigkeit und Gewissenhaftigkeit. Dabei wird u.a. eine nachhaltige Beeinflussung von gewissenhaftem Verhalten durch die kognitive Leistungsfähigkeit angenommen: Personen mit einer geringeren Gf-Ausprägung sollen über den Zeitraum mehrerer Jahre eine höhere Gewissenhaftigkeit entwickeln (Chamorro-Premuzic & Arteché, 2008; Chamorro-Premuzic & Furnham, 2004; Moutafi et al., 2004). Durch Ordnung, Sorgfalt, Zielstrebigkeit und Beständigkeit sollen die Nachteile einer niedrigeren Gf-Ausprägung kompensiert werden. Darüber hinaus soll durch diesen Kompensationsmechanismus eine Zunahme von Gc über die Zeit auftreten, da Personen, die sorgfältiger, zielstrebig und beständiger sind, über die Zeit mehr Wissen bzw. Gc erwerben (Furnham et al., 2005). Die Annahmen eines negativen Zusammenhangs zwischen Gf und G sowie eines positiven Zusammenhangs zwischen Gc und G konnten mittels Strukturanalysen, denen z.T. Längsschnittdaten zugrunde lagen in einer Arbeit von Chamorro-Premuzic und Arteché (2008) bestätigt werden (siehe Kapitel 2.3.2).

Analog zu Ackerman und Heggestad (1997) kann für den Zusammenhang zwischen g und G in beiden Analyse-Varianten – Zuordnung der Fähigkeitskonstrukte gemäß Autoren publizierter Studien vs. CHC-Theorie – ein Nulleffekt in der Population nicht ausgeschlossen werden. In Judge et al. (2007) und Poropat (2009) wurden ebenfalls sehr kleine Populationszusammenhänge zum Zusammenhang zwischen g und G berichtet. Allerdings ist in diesen Arbeiten nicht ersichtlich, ob ein Nulleffekt vorliegt (siehe oben). Ein Publikationsbias kann nach der „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien berücksichtigt wurde,

ausgeschlossen werden. Auch wenn sich in dieser Analyse-Variante die publizierten Effektgrößen symmetrisch um die mittlere Effektgröße verteilen, konnte nach der Simulationsmethode ein Publikationsbias nicht ausgeschlossen werden. Sowohl im positiven als auch im negativen Wertebereich sind mehr signifikante ( $p \leq .05$ , zweiseitig) als nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) publiziert, als man bei einer zufälligen Verteilung signifikanter ( $p \leq .05$ , zweiseitig) und nichtsignifikanter Effektgrößen ( $p > .05$ , zweiseitig) annehmen würde. Dementsprechend ist davon auszugehen, dass nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) sowohl mit positiven als auch negativen Vorzeichen häufiger nicht zur Publikation gelangt sind. Da die im Rahmen der vorliegenden Arbeit entwickelte Simulationsmethode auch eine Identifikation eines Publikationsbias erlaubt, wenn sich die publizierten Effektgrößen symmetrisch um die mittlere Effektgröße verteilen, sind die hier festgestellten Unterschiede der Ergebnisse hinsichtlich des Vorhandenseins eines Publikationsbias nicht verwunderlich.

In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, konnte sowohl nach der „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) als auch nach der Simulationsmethode ein Publikationsbias nicht ausgeschlossen werden. Dabei kann davon ausgegangen werden, dass sich die publizierten Effektgrößen asymmetrisch um die mittlere Effektgröße verteilen. Darüber hinaus wurden mehr signifikante ( $p \leq .05$ , zweiseitig) als nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) publiziert, als man bei einer zufälligen Verteilung signifikanter ( $p \leq .05$ , zweiseitig) und nichtsignifikanter Effektgrößen ( $p > .05$ , zweiseitig) erwarten würde. Dementsprechend ist nicht auszuschließen, dass Studien aufgrund nichtsignifikanter Effektgrößen zum Zusammenhang zwischen  $g$  und  $G$  häufiger nicht zur Publikation gelangt sind.

Bezüglich der Moderatoranalysen betrug die erklärte Varianz in den Studien, in denen der NEO-FFI zur Erfassung von  $G$  eingesetzt wurde, mehr als 75%. Offenbar sind die konzeptionellen Unterschiede zwischen dem NEO-FFI und anderen Persönlichkeitsinventaren in Bezug auf Gewissenhaftigkeit relevant für die Zusammenhänge mit  $g$ .

Zum Zusammenhang zwischen  $G_f$  und  $G$  kann in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren vorgenommen wurde, ein Nulleffekt in der Population nicht ausgeschlossen werden. Nach der Exklusion einer einzelnen Studie mit einer großen Stichprobe (De Fruyt et al., 2006;  $n = 20,175$ ) kann von einem negativen Populationszusammenhang ausgegangen werden. In der Analyse-Variante, in der eine Zuordnung



der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, konnte unabhängig von der Exklusion bzw. Inklusion der Studie von De Fruyt et al. (2006,  $n = 20,175$ ) ein Nulleffekt in der Population nicht ausgeschlossen werden.

Ein Publikationsbias kann nach der „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien vorgenommen wurde, ausgeschlossen werden. Trotz symmetrischer Verteilung der publizierten Effektgrößen um die mittlere Effektgröße kann nach der Simulationsmethode ein Publikationsbias nicht ausgeschlossen werden. Dabei ist davon auszugehen, dass mehr signifikante ( $p \leq .05$ , zweiseitig) als nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) publiziert wurden, als man bei einer zufälligen Verteilung signifikanter ( $p \leq .05$ , zweiseitig) und nichtsignifikanter Effektgrößen ( $p > .05$ , zweiseitig) um den Populationszusammenhang erwarten würde. Offenbar sind Studien, die nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) berichten, häufiger nicht zur Publikation gelangt als Studien, die signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) zum Zusammenhang zwischen Gf und G berichtet haben. Dieses Publikationsbias deutet auf ein Bemühen hin, die Theorie, nach der Personen mit einer geringeren Gf-Ausprägung eine höhere Gewissenhaftigkeit haben sollen (bspw. Chamorro-Premuzic & Arteche, 2008; Chamorro-Premuzic & Furnham, 2004; Moutafi et al., 2004), korrelativ zu bestätigen, wobei offenbar die Tendenz besteht, nicht dazu passende empirische Evidenz zu ignorieren bzw. zu unterdrücken. In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, kann sowohl nach der „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) als auch nach der Simulationsmethode ein Publikationsbias nicht ausgeschlossen werden. Auch in dieser Analyse-Variante kann davon ausgegangen werden, dass eine Tendenz besteht, die Theorie, dass Personen mit einer geringeren Gf-Ausprägung eine höhere Gewissenhaftigkeit haben, korrelativ zu bestätigen und dazu nicht passende empirische Evidenz zu unterdrücken. Hinsichtlich der Methoden zur Identifikation des Publikationsbias zeigt die Simulationsmethode konsistent – unabhängig von einer symmetrischen Verteilung der publizierten Effektgrößen – ein Publikationsbias im Vergleich zur „trim & fill“-Methode (Duval, 2005; Duval & Tweedie, 2000a, 2000b) an.

Hinsichtlich der Moderatoranalysen konnte in den Studien, in denen der NEO-PI-R zur Erfassung von G verwendet wurde, ein negativer Zusammenhang nicht ausgeschlossen werden. Hingegen konnte in den Studien, in denen der NEO-FFI verwendet wurde, ein Nulleffekt zum

Zusammenhang zwischen Gf und G nicht ausgeschlossen werden. Weiterhin gab es Unterschiede bezüglich der eingesetzten kognitiven Leistungstests, die für den Zusammenhang zwischen Gf und G relevant sind. In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien vorgenommen wurde, kann in den Studien, in denen nicht der BRT zur Erfassung von Gf verwendet wurde, ein Effekt von  $\rho = -.07$  angenommen werden. Hingegen kann in den Studien, in denen der BRT zur Erfassung von Gf verwendet wurde, ein Nulleffekt zum Zusammenhang zwischen Gf und G nicht ausgeschlossen werden. In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen und die Studie von De Fruyt et al. (2006),  $n = 20,175$ ) exkludiert wurde, konnte in den Studien, in denen die Raven-Tests verwendet wurden, von einem negativen Populationszusammenhang ausgegangen werden. In den Studien, in denen andere kogLt als die Raven-Tests verwendet wurden, konnte hingegen ein Nulleffekt zum Zusammenhang zwischen Gf und G nicht ausgeschlossen werden. Insgesamt kann aufgrund der Ergebnisse davon ausgegangen werden, dass konzeptionell bedingte Unterschiede zum Zusammenhang zwischen Gf und G in Abhängigkeit von den verwendeten Persönlichkeitsinventaren und kognitiven Leistungstests bestehen.

Zum Zusammenhang zwischen Gc und G kann in beiden Analyse-Varianten von einem positiven Populationszusammenhang in der Population ausgegangen werden. Allerdings wird dieser positive Effekt durch die Studie von Reeve et al. (2006,  $n = 71,887$ ) moderiert. Nach der Exklusion dieser Studie kann in beiden Analyse-Varianten ein Nulleffekt in der Population analog zu Ackerman und Heggestad (1997) nicht ausgeschlossen werden.

Hinsichtlich eines Publikationsbias kann in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien vorgenommen wurde, von einer asymmetrischen Verteilung der publizierten Effektgrößen um die mittlere Effektgröße ausgegangen werden („trim & fill“-Methode; Duval, 2005; Duval & Tweedie, 2000a, 2000b). Allerdings beruht die Annahme einer asymmetrischen Verteilung auf einer einzelnen Effektgröße bei einer inkludierten Studienzahl von lediglich  $k = 8$ . Dementsprechend sollte eine Interpretation hinsichtlich eines Publikationsbias unter Vorbehalt vorgenommen werden. Nach der Simulationsmethode kann ein Publikationsbias ausgeschlossen werden. Aufgrund der geringen Studienanzahl sollten die Ergebnisse in dieser Analyse-Variante als nicht robust bewertet werden: Weitere Publikationen zum Zusammenhang zwischen Gc und G bleibt es in dem Fall abzuwarten. In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie vorgenommen wurde, ist sowohl nach der „trim & fill“-Methode (Duval, 2005; Duval & Tweedie,

2000a, 2000b) als auch nach der Simulationsmethode ein Publikationsbias bei einer inkludierten Stichprobenanzahl von  $k = 22$  nicht auszuschließen. Dabei kann davon ausgegangen werden, dass sich die publizierten Effektgrößen asymmetrisch um die mittlere Effektgröße verteilen und dass mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) als nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) im positiven und negativen Wertebereich publiziert wurden, als man bei einer zufälligen Verteilung von signifikanten ( $p \leq .05$ , zweiseitig) und nichtsignifikanten Effektgrößen ( $p > .05$ , zweiseitig) – ohne Publikationsbias – erwarten würde. Offenbar kann in dieser Analyse-Variante davon ausgegangen werden, dass nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) zum Zusammenhang zwischen  $G_c$  und  $G$  häufiger nicht zur Publikation gelangt sind.

Mit den in der vorliegenden Arbeit gerechneten Moderatoren „Persönlichkeitsinventar“ und „kognitiver Leistungstest“ wurden zum Zusammenhang zwischen  $G_c$  und  $G$  ausnahmslos weniger als 75% der Varianz erklärt. Deshalb kann nicht ausgeschlossen werden, dass weitere Moderatoren existieren, die den Zusammenhang zwischen  $G_c$  und  $G$  erklären.

Insgesamt stehen die Ergebnisse zum Zusammenhang zwischen  $g$ ,  $G_f$ ,  $G_c$  und Gewissenhaftigkeit z.T. im Widerspruch zur Literatur. Die Annahme eines negativen Zusammenhangs zwischen  $G_f$  und  $G$  (vgl. Chamorro-Premuzic & Arteche, 2008; Chamorro-Premuzic & Furnham, 2004; Moutafi et al., 2004) konnte partiell und entgegen den Ergebnissen in Ackerman und Heggestad (1997) unterstützt werden. In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien berücksichtigt und die Studie von De Fruyt et al. (2006,  $n = 20,175$ ) exkludiert wurde, kann tatsächlich von einem negativen Populationszusammenhang zwischen  $G_f$  und  $G$  ausgegangen werden. Allerdings sollte in dem Fall berücksichtigt werden, dass mehr signifikante Effektgrößen ( $p \leq .05$ , zweiseitig) mit negativen Vorzeichen als nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) publiziert wurden, als bei einer zufälligen Verteilung um die Populationseffektgröße – ohne Publikationsbias – zu erwarten gewesen wäre. Dementsprechend ist ein Publikationsbias nicht auszuschließen, bei dem häufiger nichtsignifikante Effektgrößen ( $p \leq .05$ , zweiseitig) nicht zur Publikation gelangt sind. Möglicherweise ist das Ergebnis eines negativen Populationszusammenhangs auch durch das Publikationsbias bedingt.

Die metaanalytischen Ergebnisse der vorliegenden Arbeit unterstützen partiell die Annahmen bzw. Ergebnisse eines positiven Zusammenhangs zwischen  $G_c$  und  $G$  (bspw. Chamorro-Premuzic & Furnham, 2004; Chamorro-Premuzic & Arteche, 2008). Allerdings kann nach der Exklusion der

Studie von Reeve et al. (2006,  $n = 71,887$ ) ein Nulleffekt in der Population nicht mehr ausgeschlossen werden. Dieser Effekt konnte für beide Analyse-Varianten festgestellt werden.

*Zum Publikationsbias:*

Bei mindestens 15 (50%) der insgesamt 30 metaanalytisch untersuchten Zusammenhänge (fünf Persönlichkeitsdimensionen x drei Intelligenzkonstrukte x zwei Analyse-Varianten) kann ein Publikationsbias gemäß der in der vorliegenden Arbeit entwickelten Simulationsmethode nicht ausgeschlossen werden. Von 13 Populationszusammenhängen, bei denen die Null nicht im Glaubwürdigkeitsintervall liegt, kann in sechs Fällen (46.2%) ein Publikationsbias nicht ausgeschlossen werden (Tabelle 76).

**Tabelle 76**

Populationszusammenhänge von  $g$ ,  $G_f$  und  $G_c$  mit den Persönlichkeitsdimensionen  $Ne$ ,  $E$ ,  $O$ ,  $V$  und  $G$  bei denen das Glaubwürdigkeitsintervall die Null nicht mit einschließt.

|       | $Ne$<br>AVar1/AVar2             | $E$<br>AVar1/AVar2 | $O$<br>AVar1/AVar2                  | $V$<br>AVar1/AVar2 | $G$<br>AVar1/AVar2 |
|-------|---------------------------------|--------------------|-------------------------------------|--------------------|--------------------|
| $g$   | <b>-.09</b> / -.11 <sup>1</sup> | - / -              | .19 <sup>1</sup> / .20 <sup>1</sup> | - / -              | - / -              |
| $G_f$ | - / -                           | - / -              | <b>.09</b> / <b>.10</b>             | .01 / -            | <b>-.07</b> / -    |
| $G_c$ | -.07 / <b>-.07</b>              | - / -              | <b>.24</b> / .29                    | -.02 / -           | - / -              |

*Anmerkungen.* AVar1 = Analyse-Variante, in der die Zuordnung der Fähigkeitskonstrukte gemäß der Autoren der Studien erfolgte. AVar2 = Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie im Nachhinein durch den Autor der vorliegenden Arbeit vorgenommen wurde. Fett gedruckt: Populationszusammenhänge, bei denen das Glaubwürdigkeitsintervall die Null nicht mit einschließt und ein Publikationsbias nicht ausgeschlossen werden kann; <sup>1</sup> = keine schlüssige Interpretation, daher keine Aussage zum Publikationsbias mittels Simulationsmethode.

Die Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien birgt möglicherweise mehr Variabilität bzw. weniger Präzision als die Analyse-Variante, die einheitlich nach einer Intelligenzstrukturtheorie durchgeführt wird. Die Zuordnung der Tests zu den Fähigkeitskonstrukten unverändert zu lassen und nicht im Nachhinein gemäß der CHC-Theorie zuzuordnen, bietet jedoch den Vorteil, eine reale Abbildung der Uneindeutigkeiten der Zuordnung der Tests zu den Konstrukten zu berücksichtigen. Damit wurde nicht zuletzt auch versucht, ein sozialkonstruktivistisches Wissenschaftsverständnis in die vorliegende Arbeit zu integrieren: Die Zuordnung der Tests zu den Fähigkeitskonstrukten erfolgte z.T. durch unterschiedliche, von Personen erschaffenen Annahmen.

Durch die Berücksichtigung von zwei Analyse-Varianten (Zuordnung der Fähigkeitskonstrukte gemäß der Autoren vs. Zuordnung gemäß CHC-Theorie) wurde aber auch der gestellten Forderung

von Rothstein und Bushman (2012) entsprochen, das Publikationsbias mit verschiedenen Methoden zu untersuchen. Aufgrund der Ergebnisse der vorliegenden Arbeit kann nicht ausgeschlossen werden, dass zum Zusammenhang zwischen  $g$ ,  $G_f$ ,  $G_c$  und den Persönlichkeitsdimensionen des FFM in ca. 50% der untersuchten Zusammenhänge ein Publikationsbias besteht. Dabei kann davon ausgegangen werden, dass Studien, die nichtsignifikante Effektgrößen ( $p > .05$ , zweiseitig) berichten, häufiger nicht zur Publikation gelangt sind. Ob die Nichtpublikation dieser Studien primär auf die Verantwortlichen wissenschaftlicher Fachzeitschriften zurückzuführen ist, oder ob Forscher bzw. Forschergruppen nichtsignifikante Ergebnisse als weniger relevant und publizierbar einschätzten, kann im Rahmen der vorliegenden Arbeit nicht eindeutig beantwortet werden (vgl. Coursol & Wagner, 1986; Ferguson & Heene, 2012). Allerdings kann davon ausgegangen werden, dass grundsätzlich ein Bestreben besteht, signifikante Effektgrößen zu generieren.

Im Rahmen der vorliegenden Arbeit traten häufiger signifikante Effektgrößen entlang der eingezeichneten Signifikanzfunktionen auf. Hingegen traten im nichtsignifikanten Wertebereich entlang der eingezeichneten Signifikanzfunktion überzufällig weniger Effektgrößen auf. Dieser Effekt ist auch damit kompatibel, dass Effektgrößen, die zunächst in der Nähe der Signifikanzgrenze auftreten, im nachhinein durch Vergrößerung der Stichprobe in den signifikanten Bereich „gedrückt“ werden (vgl. Ferguson, 2009; Kirk, 1996). Obwohl bereits Rosnow und Rosenthal (1989, p. 1277) in ironischer Weise kritisiert haben: „...surely, God loves the .06 nearly as much as the .05“, kann auch zum gegenwärtigen Zeitpunkt nicht ausgeschlossen werden, dass Veröffentlichungsverzerrungen nahe der Signifikanzgrenze häufig aufzutreten.

### **6. 3 Grenzen der vorliegenden Arbeit und verbleibendes Arbeitsprogramm**

In der vorliegenden Arbeit wurden zwei Analyse-Varianten gerechnet. In der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien berücksichtigt wurde, führte eine nicht explizite Nennung von  $g$ ,  $G_f$  oder  $G_c$  zur Exklusion der Studie. Hingegen wurden in der Analyse-Variante, in der eine Zuordnung gemäß der CHC-Theorie (McGrew, 2009) vorgenommen wurde, die Fähigkeitskonstrukte auch bei nicht expliziter Nennung von  $g$ ,  $G_f$  oder  $G_c$  den Intelligenzkonstrukten  $g$ ,  $G_f$  oder  $G_c$  zugeordnet – soweit dies möglich war (siehe Inklusionskriterien, Kapitel 4.2). Aufgrund dessen ist die Anzahl inkludierter Studien in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der CHC-Theorie (McGrew, 2009) nach der Vorlage von Süß und Beauducel (2011) erfolgte, größer als in der Analyse-Variante, in der eine Zuordnung der Fähigkeitskonstrukte gemäß der Autoren publizierter

Studien vorgenommen wurde. Die unterschiedliche Anzahl inkludierter Studien in Abhängigkeit von der Analyse-Variante kann hinsichtlich der Vergleichbarkeit, bei der eine Ambiguität der Interpretation nicht auszuschließen ist, als eine Einschränkung der vorliegenden Arbeit gewertet werden. Um diese Ambiguität der Interpretation zu vermeiden, könnten in zukünftigen Untersuchungen nur die Studien inkludiert werden, in denen eine explizite Zuordnung bzw. Nennung der Fähigkeitskonstrukte gemäß der Autoren publizierter Studien erfolgte. Anschließend könnten zwei Analyse-Varianten gerechnet werden, bei der die inkludierte Studienzahl nicht verändert werden würde: In der ersten Analyse-Variante würde die Zuordnung der Fähigkeitskonstrukte gemäß Autoren der publizierten Studien berücksichtigt werden und in der zweiten Analyse-Variante würde eine Zuordnung der Fähigkeitskonstrukte im Nachhinein nach einem einheitlichen Intelligenzstrukturmodell erfolgen. Auch wenn durch dieses Vorgehen insgesamt weniger Studien inkludiert werden würden, wäre eine bessere Vergleichbarkeit der metaanalytischen Kennwerte der zwei Analyse-Varianten hinsichtlich der Auswirkungen unterschiedlicher Zuordnungen der Fähigkeitskonstrukte gegeben.

Die Höhe der Populationszusammenhänge zwischen  $g$ ,  $G_f$ ,  $G_c$  und dem FFM können aufgrund der Ergebnisse der vorliegenden Arbeit als gering eingeschätzt werden. Allerdings kann die Frage nicht beantwortet werden, ob der Zusammenhang zwischen Intelligenz und Persönlichkeit generell gering ist oder ob zur Untersuchung des Zusammenhangs zwischen Intelligenz und Persönlichkeit auch die Generalitätsebene der Fähigkeitskonstrukte bzw. Persönlichkeitsdimensionen eine entscheidende Bedeutung hinsichtlich der Höhe des Zusammenhangs einnimmt. Einen Hinweis, dass eine niedrigere Generalitätsebene hinsichtlich der Fähigkeitskonstrukte von Relevanz sein könnte, zeigte sich bspw. zum Zusammenhang zwischen  $G_f$  und  $Ne$  in der vorliegenden Arbeit: Die Korrelationshöhe zwischen  $G_f$  und  $Ne$  variierte in Abhängigkeit des Anteils der Gewichtung numerischer Fähigkeiten (siehe Abbildung 16, S. 78; Abbildung 17, S. 80; vgl. Beauducel et al., 2007). Dementsprechend kann die Berücksichtigung spezifischerer Fähigkeitskonstrukte bzw. Persönlichkeitsdimensionen auf einer niedrigeren Generalitätsebene zum Zusammenhang zwischen Intelligenz und Persönlichkeit von Relevanz sein und sollte in zukünftigen Arbeiten berücksichtigt werden (vgl. Wolf & Ackerman, 2005).

Da zum Zusammenhang zwischen Intelligenz und Persönlichkeit die altersspezifische Entwicklung von Personen eine entscheidende Rolle einzunehmen scheint (Ackerman, 1996; Ashton et al., 2000; Chamorro-Premuzic & Arteché, 2008; Chamorro-Premuzic & Furnham, 2004; DeYoung et al., 2005, Moutafi et al., 2004; Raine et al., 2002; Ziegler et al., 2012), könnte die

Berücksichtigung des Lebensalters der Studienteilnehmer in Jahren sinnvoll sein. In der vorliegenden Arbeit konnte bspw. zum Zusammenhang zwischen der kognitiven Leistungsfähigkeit und der Persönlichkeitsdimension Offenheit für Erfahrungen mit keinem Moderator mehr als 75% der Varianz erklärt werden. Um die bestehenden, aber in der vorliegenden Arbeit nicht identifizierten Subpopulationen zu untersuchen, könnte es hilfreich sein, eine Moderatoranalyse hinsichtlich des „Lebensalters der Studienteilnehmer“ durchzuführen (vgl. Wolf & Ackerman, 2005). Selbstverständlich ist sich der Autor der vorliegenden Arbeit bewusst, dass korrelationsbasierte metaanalytische Ergebnisse, denen Querschnittsdesigns zu Grunde liegen, nicht optimal geeignet sind, altersspezifische Entwicklungen bzw. Investmenttheorien zum Zusammenhang zwischen Intelligenz und Persönlichkeit zu untersuchen. Da jedoch in der Forschungspraxis davon auszugehen ist, dass häufig nur Datensätze mit Querschnittsdesign für metaanalytische Untersuchungen zur Verfügung stehen, sollten in dem Fall Abstriche zugunsten des Erkenntnisgewinns in Kauf genommen werden.

Hinsichtlich des Publikationsbias kann als Grenze der vorliegenden Arbeit die Nichtberücksichtigung nicht publizierter Studien in wissenschaftlichen Fachzeitschriften aufgeführt werden. Da davon auszugehen ist, dass Studien mit signifikanten Effektstärken bevorzugt publiziert werden und Studien mit nichtsignifikanten Effektstärken häufig nicht zur Publikation gelangen (Ferguson & Brannick, 2012), ist nicht auszuschließen, dass die Höhe des Publikationsbias in der vorliegenden Arbeit überschätzt ist. Allerdings wird die Frage, ob nichtpublizierte Studien in Metaanalysen berücksichtigt werden sollten, kontrovers diskutiert (Ferguson & Brannick, 2012; Rothstein & Bushman, 2012). Um den Einfluss publizierter Studien vs. nicht publizierter Studien zu untersuchen, wäre es in zukünftigen metaanalytischen Untersuchungen sinnvoll, sowohl publizierte als auch nichtpublizierte Studien zu inkludieren. Mittels Moderatoranalyse könnten die Effektgrößen publizierter vs. nichtpublizierter Studien verglichen werden. Dabei wäre zu erwarten, dass bei den inkludierten Studien, die in Fachzeitschriften publiziert sind, eine größere mittlere Effektgröße auftritt als bei den inkludierten Studien, die nicht publiziert sind. Darüber hinaus berichten Rothstein und Bushman (2012), dass in höherrangigen Fachzeitschriften die Höhe des Zusammenhangsmaßes von entscheidender Bedeutung sein kann, um eine Studie zu publizieren. Dementsprechend könnte es ebenfalls empfehlenswert sein, den „impact factor“ einer wissenschaftlichen Zeitschrift als Moderator in metaanalytischen Untersuchungen zu berücksichtigen. Dabei würde man eine größere Effektgröße für die berücksichtigten Studien mit größerem „impact factor“ erwarten als für die inkludierten Studien mit niedrigerem „impact factor“.

Gemäß der im Rahmen der vorliegenden Arbeit entwickelten Simulationsmethode konnten zum Zusammenhang zwischen  $g$  und  $N_e$  sowie  $g$  und  $O$  keine schlüssigen Interpretationen hinsichtlich eines Publikationsbias angeboten werden. Es ist davon auszugehen, dass es insbesondere dann zu uneindeutigen Interpretationen kommen kann, wenn große bzw. heterogene Stichproben inkludiert sind. Um die Präzision hinsichtlich der Identifikation des Publikationsbias auch in den Fällen, in denen sehr große und/oder heterogene Stichproben auftreten, zu gewährleisten, sollte zukünftig eine Gewichtung der tatsächlich publizierten und gezogenen Effektgrößen anhand der Stichprobengröße bei der Ermittlung des Index berücksichtigt werden.

Letztlich sollten im Sinne eines umfassenderen Verständnisses der Konstruktvalidität (vgl. Cronbach & Meehl, 1955) auch weitere Untersuchungsmethoden zum Zusammenhang zwischen Intelligenz und Persönlichkeit berücksichtigt werden. Eine Möglichkeit, den Zusammenhang zwischen Intelligenz und Persönlichkeit zu untersuchen, wäre die Berücksichtigung biopsychologischer Variablen. Bspw. sind bereits verschiedene Studien publiziert worden, die individuelle Unterschiede in der Intelligenzleistung und/oder Persönlichkeit mit dem Elektroenzephalogramm (EEG) untersuchen (bspw. Beauducel, Brocke, & Leue, 2006; Jaušovec, 2000; Lutzenberger, Birbaumer, Flor, Rockstroh & Elbert, 1992; Razoumnikova, 2003; Thatcher, North & Biver, 2008). Die Kombination von psychometrischen Persönlichkeitstests und kognitiven Leistungstests mit biopsychologischen Variablen könnte vielversprechend sein, um weitere, detaillierte Erkenntnisse zum Zusammenhang zwischen Intelligenz und Persönlichkeit zu generieren.



## Literatur

In die Metaanalyse inkludierte Studien sind mit einem Stern (\*) markiert.

- Abraham, J. D. & Morrison, J. D., Jr. (2002). *Performance perspectives inventory: PPI technical manual, version b1.3*. Tulsa, OK: A & M Psychometrics, LLC.
- Ackerman, P. L. (1994). Intelligence, attention, and learning: Maximal and typical performance. In D. K. Detterman (Ed.), *Current topics in human intelligence. Vol. 4: Theories of intelligence* (pp. 1-27). Norwood, NJ: Ablex.
- Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests and knowledge. *Intelligence*, 22, 227-257.
- \*Ackerman, P. L. (2000). Domain-Specific knowledge as the “Dark Matter” of adult intelligence: Gf/Gc, personality and interest correlates. *The Journal of Gerontology*, 55B, 69-84.
- \*Ackerman, P. L., Bowen, K. R., Beier, M. E. & Kanfer, R. (2001). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology*, 93, 797-825.
- Ackerman, P. L. & Goff, M. (1994). Typical intellectual engagement and personality: Reply to Rocklin (1994). *Journal of Educational Psychology*, 86, 150-153.
- Ackerman, P. L. & Heggestad, E. D. (1997). Intelligence, personality and interests: evidence for overlapping traits. *Psychological Bulletin*, 121, 219-245.
- \*Ackerman, P. L., Kanfer, R. & Goff, M. (1995). Cognitive and Noncognitive Determinants and Consequences of Complex Skill Acquisition. *Journal of Experimental Psychology*, 1, 270-304.
- \*Allik, J. & Realo, A. (1997). Intelligence, academic abilities, and personality. *Personality and Individual Differences*, 23, 809-814.
- Allport, G. W. & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, 47, No. 211.
- \*Allworth, E. & Hesketh, B. (2000). Job requirements biodata as a predictor of performance in customer service roles. *International Journal of Selection and Assessment*, 8, 137-147.
- Amelang, M., Bartussek, D., Stemmler, G. & Hagemann, D. (2006). *Differentielle Psychologie und Persönlichkeitsforschung* (6<sup>th</sup> ed.): Stuttgart: W. Kohlhammer Verlag.
- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R* (2<sup>nd</sup> ed.). Göttingen: Hogrefe.
- Andresen, B. (2001). Konzepte und Fragebogenskalen zur Einordnung von Psychosetendenzen in die differentiell-psychologischen Faktoren der Persönlichkeit. In: B. Andresen & R. Maß

## Literatur

- (Eds.), *Schizotypie - Psychometrische Entwicklungen und piopsychologische Forschungsansätze* (pp. 4-43). Göttingen: Hogrefe.
- Arteche, A., Chamorro-Premuzic, T., Ackerman, P. & Furnham, A. (2009). Typical intellectual engagement as a byproduct of openness, learning approaches, and self-assessed intelligence. *Educational Psychology: An International Journal of Experimental Educational Psychology*, 29, 357-367.
- \*Arteche, A., Chamorro-Premuzic, T., Furnham, A. & Crump, J. (2008). The relationship of trait EI with personality, IQ and sex in a UK Ssmple of employees. *International Journal of Selection and Assessment*, 16, 422-427.
- \*Ashton, M. C., Lee, K., Vernon, P. A. & Jang, K. (2000). Fluid intelligence, crystallized intelligence, and the openness/intellect factor. *Journal of Research in Personality*, 34, 198-207.
- \*Atwater, L. E. (1997). Beyond cognitive ability: Improving the prediction of performance. *Journal of Business and Psychology*, 7, 27-44.
- \*Austin, E. J., Deary, I. J. & Gibson, G. J. (1997). Relationships between ability and personality: Three hypotheses tested. *Intelligence*, 25, 49-70.
- \*Austin, E. J., Deary, I. J., Whiteman, M. C., Fowkes, F. G. R., Pedersen, N. L., Rabbitt, P., Bent, N. & McInnes, L. (2002). Relationships between ability and personality: Does Intelligence contribute positively to personal and social adjustment? *Personality and Individual Differences*, 32, 1391-1411.
- Austin, E. J., Deary, I. J. & Willock, J. (2001). Personality and intelligence as predictors of economic behaviour in Scottish farmers. *European Journal of Personality*, 15, 123-137.
- Baddeley, A. D. (1968). A 3 min reasoning test based on grammatical transformation. *Psychonomic Science*, 10, 341-342.
- Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology*, 23, 611-626.
- Baltes, P. B. (1997). On the incomplete architecture of human ontogeny: Selection, optimization, and compensation as foundation of developmental theory. *American Psychologist*, 52, 366-380.
- Barrick, M. R. & Mount, M. K. (1995). *The Personal Characteristics Inventory manual*. Unpublished manuscript, University of Iowa, Iowa City.
- Barrick, M. R., Mount, M. K. & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessement*, 9, 9-30.

## Literatur

- \*Barrick, M. R., Stewart, G. L., Neubert, M. J. & Mount, K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, 83, 377-391.
- Barrick, M. R., Stewart, G. L. & Piotrowski, M. (2002). Personality and job performance: Test of the mediating effects of motivation among sales representatives. *Journal of Applied Psychology*, 87, 43-51.
- \*Bartone, P. T., Snook, A. S. & Tremble, T. R. (2002). Cognitive and personality predictors of leader performance in West Point Cadets. *Military Psychology*, 14, 321-338.
- \*Bates, T. C. & Shieles, A. (2003). Crystallized intelligence as a product of speed and drive for experience: The relationship of inspection time and openness to g and Gc. *Intelligence*, 31, 275-287.
- \*Batey, M., Chamorro-Premuzic, T. & Furnham, A. (2010). Individual differences in ideational behavior: Can the Big Five and psychometric intelligence predict creativity scores? *Creativity Research Journal*, 22, 90-97.
- Beauducel, A. (2001a). On the generalizability of factors: The influence of changing contexts of variables on different methods of factor extraction. *Methods of Psychological Research*, 6, 1-24.
- Beauducel, A. (2001b). Problems with parallel analysis in data sets with oblique simple structure. *Methods of Psychological Research*, 6, 141-157.
- Beauducel, A., Brocke, B. & Leue, A. (2006). Energetical bases of extraversion: Effort, arousal, EEG, and performance. *International Journal of Psychophysiology*, 62, 212-223.
- Beauducel, A., Brocke, B. & Liepmann, D. (2001). Perspectives on fluid and crystallized intelligence: Facets for verbal, numerical, and figural intelligence. *Personality and Individual Differences*, 30, 977-994.
- \*Beauducel, A., Liepmann, D., Felfe, J. & Nettelstroth, W. (2007): The impact of different measurement models for fluid and crystallized intelligence on the correlation with personality traits. *European Journal of Psychological Assessment*, 23, 71-78.
- Begg, C. B. (1994). Publication bias. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399-409). New York: Russell Sage.
- Begg, C. B. & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50, 1088-1101.
- \*Beier, M. A. & Ackerman, P. L. (2001). Current-events knowledge in adults: An investigation of age, intelligence, and nonability determinants. *Psychology and Aging*, 16, 615-628.

## Literatur

- \*Beier, M. A., Cambell, M. & Crook, A. E. (2010). Developing and demonstrating knowledge: Ability and non-ability determinants of learning and performance. *Intelligence*, 38, 179-186.
- \*Biderman, M. D., Nguyen, N. T. & Sebren, J. (2008). Time-on-task mediates the conscientiousness-performance relationship. *Personality and Individual Differences*, 44, 887-897.
- \*Bipp, T. & Kleinbeck, U. (2011). The effect of Neuroticism in the process of goal pursuit. *Personality and Individual Differences*, 51, 454-459.
- \*Bipp, T., Steinmayr, R. & Spinath, B. (2008). Personality and achievement motivation: Relationship among Big Five domain and facet scales, achievement goals, and intelligence. *Personality and Individual Differences*, 44, 1454-1464.
- \*Black, J. (2000). Personality testing and Police selection: Utility of the 'Big Five'. *New Zealand Journal of Psychology*, 29, 2-9.
- Blickle, G. (1996) Personality traits, learning strategies, and performance. *European Journal of Personality*, 10, 337-352.
- Blickle, G., Wendel, S. & Ferris, G. R. (2010). Political skills as moderator of personality – job performance relationships in socioanalytic theory: Test of the getting ahead motive in automobile sales. *Journal of Vocational Behavior*, 76, 326-335.
- Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, 117, 187-215.
- Bolger, N. & Schilling, E. A. (1991). Personality and the problems of everyday life: The role of neuroticism in exposure and reactivity to daily stressors. *Journal of Personality*, 59, 355-386.
- Borkenau, P. & Ostendorf, F. (1993). *NEO-Fünf-Faktoren-Inventar (NEO-FFI) nach Costa und McCrae. Handanweisung*. Göttingen: Hogrefe.
- \*Boudreau, J. W., Boswell, W. R., Judge, T. A. & Bretz, R. D., Jr. (2001). Personality and cognitive ability as predictors of job search among employed managers. *Personnel Psychology*, 54, 25-50.
- Brand, C. (1994). Open to experience-closed to intelligence: Why the Big Five are really the Comprehensive Six. *European Journal of Personality*, 8, 299-310.
- \*Bratko, D., Chamorro-Premuzic, T. & Saks, Z. (2006). Personality and school performance: Incremental validity of self- and peer-ratings over intelligence. *Personality and Individual Differences*, 41, 131-142.
- Brecht, B. (1963). *Leben des Galilei*. Frankfurt am Main: Suhrkamp.

## Literatur

- Brocke, B. (2000). Das bemerkenswerte Comeback der Differentiellen Psychologie: Glückwünsche und Warnungen vor einem neuen Desaster. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 21, 5-30.
- Brocke, B. & Beauducel, A. (2001). Intelligenz als Konstrukt. In E. Stern & J. Guthke (Eds.), *Perspektiven der Intelligenzforschung* (pp 13-42). Lengerich: Pabst Science Publisher.
- Brody, N. & Ehrlichman, H. (1998). *Personality Psychology. The Science of Individuality*. Prentice Hall, Inc.
- \*Busato, V. V., Frans, J. P., Elshout, J. J. & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences*, 29, 1057-1068.
- \*Carless, S. A. (1999). Career assessment: Holland's vocational interests, personality characteristics, and abilities. *Journal of Career Assessment*, 7, 125-144.
- Carroll, J. B. (1982). The measurement of intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 29-120). New York: Academic Press.
- Carroll, J. B. (1993). *Human cognitive ability. A survey of factor-analytic studies*. New York: Cambridge University Press.
- Carroll, J. B. (2005). The three-stratum theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, test, amd issues* (2<sup>nd</sup> ed., pp. 69-76). New York: Guilford.
- \*Carson, A. D., Stalikas, A. & Bizot, E. B. (1997). Correlations between the Myers-Briggs Type Indicator and Measures of Aptitudes. *Journal of Career Assessment*, 5, 81-104.
- Caruso, J. C. (2000). Reliability Generalization of the NEO Personality Scales. *Educational and Psychological Measurement*, 60, 236-254.
- Caspi, A. (2000). The child is father of the man: Personality continuities from childhood to adulthood. *Journal of Personality and Social Psychology*, 78, 158-172.
- Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *Journal of Abnormal and Scocial Psychology*, 38, 476-506.
- Cattell, R. B. (1945). The principle trait clusters for describing personality. *Psychological Bulletin*, 42, 129-161.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: a critical experiment. *Journal of Educational Psychology*, 54, 1-22.
- Cattell, R. B. (1965): *The Scientific Analysis of Personality*. Baltimore: Penguin Books.
- Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. New York: Houghton Mifflin.

## Literatur

- Cattell, R. B. (1973). *Measuring intelligence with the culture fair tests*. Champaign, IL: The Institute for Personality and Ability Testing.
- Cattell, R. B. (1987). *Intelligence: Its Structure, Growth and Action*. Amsterdam: Elsevier Science Publishers B.V.
- \*Chamorro-Premuzic, T. & Arteche, A. (2008). Intellectual competence and academic performance: Preliminary validation of a model. *Intelligence*, 36, 564-573.
- Chamorro-Premuzic, T. & Furnham, A. (2004). A possible model for understanding the personality-intelligence interface. *British Journal of Psychology*, 95, 249-264.
- \*Chamorro-Premuzic, T. & Furnham, A. (2008). Personality, intelligence and approaches to learning as predictors of academic performance. *Personality and Individual Differences*, 44, 1596-1603.
- \*Chamorro-Premuzic, T., Furnham, A. & Ackerman, P. L. (2006a). Ability and personality correlates of general knowledge. *Personality and Individual Differences*, 41, 419-429.
- \*Chamorro-Premuzic, T., Furnham, A. & Ackerman, P. L. (2006b). Incremental validity of the Typical Intellectual Engagement Scale as predictor of different academic performance measures. *Journal of Personality Assessment*, 87, 261-268.
- \*Chamorro-Premuzic, T., Furnham, A. & Moutafi, J. (2004). The relationship between estimated and psychometric personality and intelligence scores. *Journal of Research in Personality*, 38, 505-513.
- \*Chamorro-Premuzic, T., Furnham, A. & Petrides, K. (2006). Personality and intelligence: The relationship of Eysenck's Giant Three with verbal and numerical ability. *Journal of Individual Differences*, 27, 147-150.
- \*Chamorro-Premuzic, T., Moutafi, J. & Furnham, A. (2005). The relationship between personality traits, subjectively-assessed and fluid intelligence. *Personality and Individual Differences*, 38, 1517-1528.
- Chamorro-Premuzic, T., Stumm, S. & Furnham, A. (2011). *The Wiley-Blackwell Handbook of Individual Differences*. West Sussex: Wiley-Blackwell.
- \*Chen, H.-G. & Vecchio, R. P. (1992). Nested IF-THEN-ELSE constructs in end-user computing: personality and aptitude as predictors of programming ability. *International Journal of Man-Machine Studies*, 36, 843-859.
- \*Christopher, A. N., Furnham, A., Batey, M., Martin, G. N., Koenig, C. S. & Doty, K. (2010). Protetsant ethic endorsement, personality, and general intelligence. *Learning and Individual Differences*, 20, 46-50.

## Literatur

- \*Colom, R. & Quiroga, Á. (2009). Neuroticism, intelligence, and intra-individual variability in elementary cognitive tasks: Testing the mental noise hypothesis. *Psicothema, 21*, 403-408.
- Cooper, H. (1997). Some finer points in meta-analysis. In M. Hunt (Ed.), *How Science Takes Stock: The Story of Meta-analysis* (pp. 169-181). New York: Russell Sage Foundation.
- Costa, P. T. & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T. & McCrae, R. R. (1992). *Revised NEO Personality Inventory and Five Factor Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.
- Coursol, A. & Wagner, E. E. (1986). Effect of positive findings on submission and acceptance rates: A note on meta-analysis bias. *Professional Psychology: Research and Practice, 17*, 136-137.
- Cronbach, L. J. (1949). *Essentials of Psychological Testing*. New York: Halpern.
- Cronbach, L. J. & Meehl, P. L. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.
- \*De Fruyt, F., Aluja, A., Garcia, L. F., Rolland, J. P. & Jung, S. C. (2006). Positive presentation management and intelligence and the personality differentiation by intelligence hypothesis in job applicants. *International Journal of Selection and Assessment, 14*, 101-112.
- De Fruyt, F. & Mervielde, I. (1996). Personality and interests as predictors of educational streaming and achievement. *European Journal of Personality, 10*, 405-425.
- Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R. & Starr, J. M. (2000). The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish mental survey. *Intelligence, 28*, 49-55.
- \*Deuling, J. K., Denissen, J. J. A., van Zalk, M., Meeus, W. & van Aken, M. (2011). Perceived influence in groups over time: How associations with personality and cognitive ability can change over time. *Journal of Research in Personality, 45*, 576-585.
- \*DeYoung, C. G., Peterson, J. B. & Higgins, D. M. (2005). Sources of openness/intellect: Cognitive and neuropsychological correlates of the fifth factor of personality. *Journal of Psychology, 73*, 825-858.
- \*DeYoung, C. G., Shamosh, N. A., Green, A. E., Braver, T. S. & Gray, J. R. (2009). Intellect as distinct from openness: differences revealed by fMRI of working memory. *Journal of Personality and Social Psychology, 97*, 883-892.
- \*Di Fabio, A. & Busoni, L. (2007). Fluid intelligence, personality traits and scholastic success: Empirical evidence in a sample of Italian high school students. *Personality and Individual Differences, 43*, 2095-2104.

## Literatur

- Digman, J. M. (1990). Personality Structure: Emergence of the five factor model. *Annual Review of Psychology*, 41, 417-440.
- Digman, J. M. (1996). The curious history of the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 1-20). New York: Guilford Press.
- \*Djapo, N., Kolenovic-Djapo, J., Djokic, R. & Fako, I. (2011). Relationship between Cattell's 16PF and fluid and crystallized intelligence. *Personality and Individual Differences*, 51, 63-67.
- \*Dobson, P. (2000). An investigation into the relationship between neuroticism, extraversion and cognitive test performance in selection. *International Journal of Selection and Assessment*, 8, 99-109.
- \*Doerfler, T. & Hornke, L. F. (2010). Working style and extraversion: New insights into the nature of item response latencies in computer-based intelligence testing. *Journal of Research in Personality*, 44, 159-162.
- \*Dollinger, S. J. & LaMartina, A. K. (1998). A note on moral reasoning and the five-factor model. *Journal of Social Behavior and Personality*, 13, 349-358.
- Duval, S. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias in meta analysis: prevention, assessment and adjustments* (pp. 127-144). Chichester: Wiley.
- Duval, S. & Tweedie, R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta analysis. *Journal of the American Statistical Association*, 95, 89-98.
- Duval, S. & Tweedie, R. (2000b). Trim and Fill: A simple funnel-plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455-463.
- Dye, D. A., Reck, M. & McDaniel, M. A. (1993). The validity of job knowledge measures. *International Journal of Selection and Assessment*, 1, 153-157.
- \*Egan, V. (1989). Notes and shorter communications: Links between personality, ability and attitudes in a low-IQ sample. *Personality and Individual Differences*, 10, 997-1001.
- \*Egan, V. (1994). Intelligence, inspection time and cognitive strategies. *British Journal of Psychology*, 85, 305-315.
- Elshout, J. J. & Akkerman, A. E. (1975). *Vijf Persoonlijkheids-factoren test 5PFT*. Nijmegen: Berkhout BV.
- \*Ettinger, U. & Corr, P. J. (2001). The frequency accrual speed test (FAST): Psychometric intelligence and personality correlates. *European Journal of Personality*, 15, 143-152.



## Literatur

- Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, IL: Charles C. Thomas.
- Eysenck, H. J. (1970). *The structure of human personality* (3<sup>rd</sup> ed). London: Methuen.
- Eysenck, H. J. (1994). Personality and intelligence: Psychometric and experimental approaches. In R. Sternberg & P. Ruzgis (Eds.), *Personality and intelligence* (pp. 3-31). New York: Cambridge University Press.
- Eysenck, H. J. & Cookson, D. (1969). Personality in primary school children: 1. Ability and achievement. *British Journal of Educational Psychology*, 39, 109-122.
- Eysenck, H. J. & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire (adult and junior)*. London: Hodder & Stoughton.
- Eysenck, H. J. & Eysenck, M. W. (1985). *Personality and individual differences*. New York: Plenum.
- Eysenck, H. & Wilson, G. (1991). *The Eysenck Personality Profiler* (1<sup>st</sup> ed.). Guildford: Psi-Press.
- \*Farsides, T. & Woodfield, R. (2003). Individual differences and undergraduate academic success: the roles of personality, intelligence, and application. *Personality and Individual Differences*, 34, 1225-1243.
- Faul, F. & Erdfelder, E. (1992). *GPOWER: A priori, and compute sample size for MS-DOS* [Computer Program]. Bonn: Bonn University.
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40, 532-538.
- Ferguson, C. J. & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, 17, 120-128.
- Ferguson, C. J. & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, 11, 555-561.
- Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, 6, 161-180.
- \*Fine, S. (2007). Overqualification and selection in leadership training. *Journal of Leadership & Organizational Studies*, 14, 61-68.
- Fisher, R. A. (1918). The correlation between relatives on the subposition of Mendalin inheritance. *Transactions of the Royal Society of Edinburgh*, 52, 399-433.
- Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal and Social Psychology*, 44, 329-344.

## Literatur

- \*Francis, L. J. (1997). Personal and social correlates of the 'closed mind' among 16 year old adolescents in England. *Educational Studies*, 23, 429-437.
- \*Freund, P. A. & Holling, H. (2011). Who wants to take an intelligence test? Personality and achievement motivation in the context of ability testing. *Personality and Individual Differences*, 50, 723-728.
- \*Friborg, O., Barlaug, D., Martinussen, M., Rosenvinge, J. H. & Hjemdal, O. (2005). Resilience in relation to personality and intelligence. *International Journal of Methods in Psychiatric Research*, 14, 29-42.
- Funder, D.C. (2001). *The Personality Puzzle* (2<sup>nd</sup> ed.). New York: Norton.
- \*Furnham, A. (2005). Self-estimated intelligence, psychometric intelligence and personality. *Psychologica*, 48, 182-192.
- \*Furnham, A., Batey, M., Anand, K. & Manfield, J. (2008). Personality, hypomania, intelligence and creativity. *Personality and Individual Differences*, 44, 1060-1069.
- \*Furnham, A. & Chamorro-Premuzic, T. (2004a). Personality, intelligence, and art. *Personality and Individual Differences*, 36, 705-715.
- \*Furnham, A. & Chamorro-Premuzic, T. (2004b). Estimating one's own personality and intelligence scores. *British Journal of Psychology*, 95, 149-160.
- \*Furnham, A. & Chamorro-Premuzic, T. (2006). Personality, intelligence and general knowledge. *Learning and Individual Differences*, 16, 79-90.
- \*Furnham, A. & Dissou, G., Sloan, P. & Chamorro-Premuzic, T. (2007). Personality and intelligence in business people: A study of two personality and two intelligence measures. *Journal of Business and Psychology*, 22, 99-109.
- \*Furnham, A., Jensen, T. & Crump, J. (2008). Personality, intelligence and assessment centre expert ratings. *International Journal of Selection and Assessment*, 16, 356-365.
- \*Furnham, A. & Monsen, J. (2009). Personality traits and intelligence predict academic school grades. *Learning and Individual Differences*, 19, 28-33.
- \*Furnham, A., Monsen, J. & Ahmetoglu, G. (2009). Typical intellectual engagement, Big Five personality traits, approaches to learning and cognitive ability predictors of academic performance. *British Journal of Educational Psychology*, 79, 769-782.
- \*Furnham, A., Moutafi, J. & Chamorro-Premuzic, T. (2005). Personality and intelligence: Gender, the big five, self-estimated and psychometric intelligence. *International Journal of Selection and Assessment*, 13, 11-24.
- \*Furnham, A., Moutafi, J. & Paltiel, L. (2005). Intelligence in relation to Jung's personality types. *Individual Differences Research*, 3, 2-13.

## Literatur

- \*Furnham, A., Swami, V., Arteche, A. & Chamorro-Premuzic (2008). Cognitive ability, learning approaches and personality correlates of general knowledge. *Educational Psychology*, 28, 427-437.
- \*Furnham, A., Taylor, J. & Chamorro-Premuzic, T. (2008). Personality and intelligence correlates of assessment center exercises. *Individual Differences Research*, 6, 181-192.
- \*Furnham, A., Zhang, J. & Chamorro-Premuzic, T. (2006). The relationship between psychometric and self-estimated intelligence, creativity, personality and academic achievement. *Imagination, Cognition and Personality*, 25, 119-145.
- \*Gaygisiz, E. (2009). Economic and cultural correlates of road-traffic accident fatality rates in OECD countries. *Perceptual and Motor Skills*, 109, 531-545.
- \*Gentry, T. A., Polzine, K. M. & Wakefield, J. A. (1985). Human genetic markers associated with variation in intellectual abilities and personality. *Personality and Individual Differences*, 6, 111-113.
- \*George, D., Stickle, K. & Rachid, F. & Wopnford, A. (2007). The association between types of music enjoyed and cognitive, behavioral, and personality factors of those who listen. *Psychomusicology*, 19, 32-56.
- \*Gignac, G. E., Stough, C. & Loukomites, S. (2004). Openness, intelligence, and self-report intelligence. *Intelligence*, 32, 133-143.
- Glass, G. V. & Kliegl, R. M. (1983). An apology for research integration in the study of psychotherapy. *Journal of Consulting and Clinical Psychology*, 51, 28-41.
- \*Goff, M. & Ackerman, P. L. (1992). Personality – intelligence relations: Assessment of typical intellectual engagement. *Journal of Educational Psychology*, 84, 437- 552.
- \*Gold, D. P., Andres, D., Etezadi, J., Arbuckle, T., Schwartzman, A. & Chaikelson, J. (1995). Structural Equation Model of intellectual change and continuity and predictors of intelligence in older men. *Psychology and Aging*, 10, 294-303.
- Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol.2, pp. 203-234). Hillsdale, NJ: Erlbaum.
- Goldberg, L. R. (1982). From Ace to Zombie: Some explorations in the language of personality. In C. D. Spielberger & J. N. Butcher (Eds.), *Advances in personality Assessment* (Vol. 1, pp. 203-234). Hillsdale, NY: Erlbaum.
- Goldberg, L. R. (1990). An alternative “Description of personality”: The Big-Five factor structure. *Journal of Personality and Social Psychology*, 59, 1216-1229.

## Literatur

- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*, 26-42.
- Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several Five-Factor models. In I. Mervielde, I. Deary, F. De Fruyt & F. Ostendorf (Eds.), *Personality Psychology in Europe* (Vol. 7, pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.
- \*Gow, A. J., Whiteman, M. C., Pattie, A. & Deary, I. J. (2005). The personality-intelligence interface: insights from an ageing cohort. *Personality and Individual Differences, 39*, 751-761.
- \*Gregory, T., Nettelbeck, T. & Wilson, C. (2010). Openness to experience, intelligence, and successful ageing. *Personality and Individual Differences, 48*, 895-899.
- Gross, J. J., Sutton, S. K. & Ketelaar, T. (1998). Relations between affect and personality: Support for the affect-level and affective reactivity views. *Personality and Social Psychology Bulletin, 24*, 279-288.
- \*Guerin, D. W., Oliver, P. H., Gottfried, A. W., Gottfried, A. E., Reichard, R. J. & Riggio, R. E. (2011). Childhood and adolescent antecedents of social skills and leadership potential in adulthood: Temperamental approach/withdrawal and extraversion. *The Leadership Quarterly, 22*, 482-494.
- \*Gully, S. M., Payne, S. C., Koles, K. L. K. & Whiteman, J.-A. K. (2002). The impact of error training and individual differences on training outcomes: An attribute-treatment interaction perspective. *Journal of Applied Psychology, 87*, 143-155.
- \*Hambrick, D. Z., Pink, J. E., Meinz, E. J., Pettibone, J. & Oswald, F. L. (2008). The roles of ability, personality, and interests in acquiring current events knowledge: A longitudinal study. *Intelligence, 36*, 261-278.
- \*Harris, J. A. (2004). Measured intelligence, achievement, openness to experience, and creativity. *Personality and Individual Differences, 36*, 913-929.
- \*Hattrup, K., O'Connell, M. S. & Wingate, P. H. (1998). Prediction of multidimensional criteria: Distinguishing task and contextual performance. *Human Performance, 11*, 305-319.
- Härnqvist, K., Gustafsson, J. E., Muthén, B. O. & Nelson, G. (1994). Hierarchical models of ability at individual and class levels. *Intelligence, 18*, 165-187.
- Hedges, L. V. (1992). Modeling publication selection effects in meta-analysis. *Statistical Science, 7*, 246-255.
- Hedges, L. V. & Olkin, I. (1985). *Statistical methods of meta-analysis*. New York: Academic Press.

## Literatur

- Hedges, L. V. & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486-504.
- \*Hill, P. L., Turiano, N. A., Hurd, M. D., Mroczek, D. K. & Roberts, B. W. (2011). Conscientiousness and longevity: An examination of possible mediators. *Health Psychology, 30*, 536-541.
- \*Hogan, T. P., Wyckoff, L. A., Krebs, P., Jones, W. & Fitzgerald, M. P. (2004). Tolerance for error and computational estimation ability. *Psychological Reports, 94*, 1393-1403.
- Holland, J. L. (1963). Explorations of a theory of vocational choice and achievement: II. A four-year prediction study. *Psychological Reports, 12*, 547-594.
- Holland, D. C., Dollinger, S. J., Holland, C. J. & MacDonald, D. A. (1995). The relationship between psychometric intelligence and the five-factor model of personality in a rehabilitation sample. *Journal of Clinical Psychology, 51*, 79-88.
- Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2<sup>nd</sup> ed., pp. 645-685). New York: Plenum Press.
- Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 443-452). New York: Macmillan.
- Horn, J. L. (1998). A basis for research on age differences in cognitive capabilities. In J. J. McArdle & Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 57-92). Chicago, IL: Riverside.
- Horn, J. L. (2008). Spearman, g, expertise, and the nature of human cognitive capability. In P. C. Kyllonen, R. D. Roberts & L. Stankov (Eds.), *Extending intelligence: Enhancement and new constructs* (pp. 185-230). New York: Erlbaum.
- Horn, J. L. & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. I. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2<sup>nd</sup> ed., pp. 41-68). New York: Guilford.
- Horn, J. L. & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*, 253-270.
- Horn, J. L. & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26*, 107-129.
- Horn, J. L. & Noll, J. (1997). Human cognitive capabilities: Gf-Gc- Theory. In D. P. Flanagan & P. I. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53-91). New York: Guilford.

## Literatur

- Horn, J. L. & Stankov, L. (1982). Auditory and visual factors of intelligence. *Intelligence*, 6, 165-185.
- \*Hultsch, D. F., Hertzog, C., Small, B. J. & Dixon, R. A. (1999). Use it or lose it: Engaged lifestyle as a buffer of cognitive decline in aging? *Psychology and Aging*, 14, 245-263.
- Hunter, J. E. & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment*, 8, 275-292.
- Hunter, J. E. & Schmidt, F. L. (2004). *Methods of meta-analysis*. London: SAGE Publications.
- \*Hunthausen, J. M., Truxillo, D. M., Bauer, T. N. & Hammer, L. B. (2003). A field study of frame-of-reference effects on personality test validity. *Journal of Applied Psychology*, 88, 545-551.
- Hurtz, G. M. & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology*, 85, 869-879.
- \*Imai, L. & Gelfand, M. J. (2010). The culturally intelligent negotiator: The impact of cultural intelligence (CQ) on negotiation sequences and outcomes. *Organizational Behavior and Human Decision Processes*, 112, 83-98.
- Irwing, P., Cammock, T. & Lynn, R. (2001). Some evidence for the existence of a general factor of semantic memory and its components. *Personality and Individual Differences*, 30, 857-871.
- Iyengar, S. & Greenhouse, J. (1988). Selection models and the file drawer problem. *Statistical Science*, 3, 109-135.
- Jaušovec, N. (2000). Differences in cognitive processes between gifted, intelligent, creative, and average individuals while solving complex problems: An EEG Study. *Intelligence*, 28, 213-237.
- Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica*, 28, 195-226.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger/Greenwood.
- John, O. P. & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of Personality: Theory and Research* (pp. 102-138). New York/London: Guilford Press.
- Judge, T. A., Heller, D. & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology*, 87, 530-541.

## Literatur

- \*Judge, T. A., Higgins, C. A., Thoresen, C. J. & Barrick, M. R. (1999). The Big Five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, 52, 621-652.
- Judge, T. A. & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology*, 87, 797-807.
- Judge, T. A., Jackson, C. L., Shaw, J. C., Scott, B. A. & Rich, B. L. (2007). Self-efficacy and work-related performance: The integral role of individual differences. *Journal of Applied Psychology*, 92, 107-127.
- \*Kanfer, R., Ackerman, P. L. & Heggestad, E. D. (1996). Motivational skills & self-regulation for learning: A trait perspective. *Learning and Individual Differences*, 8, 185-209.
- \*Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J. & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, 116, 321-340.
- Kemery, E. R., Mossholder, K. W. & Roth, L. (1987). The power of the Schmidt and Hunter additive model of validity generalization. *Journal of Applied Psychology*, 72, 30-37.
- \*Kickul, J. & Neumann, G. (2000). Emergent leadership behaviors: The function of personality and cognitive ability in determining teamwork performance and KSAs. *Journal of Business and Psychology*, 15, 27-51.
- Kirk, R. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56, 746-759.
- Klages, L. (1926). *Zur Ausdruckslehre und Charakterkunde*. Heidelberg: Nils Kampmann Verlag.
- \*Kluge, A., Ritzmann, D. B. & Sauer, J. (2011). The interaction of drill and practise and error training with individual differences. *Cognition, Technology & Work*, 13, 103-120.
- Knuth, D. E. (1981). *The art of computer programming* (2<sup>nd</sup> ed., Vol. 2). Reading, Mass.: Addison-Wesley.
- Kramer, J. (2009). *Metaanalytische Studien zu Intelligenz und Berufsleistung in Deutschland*. Elektronisch veröffentlichte Dissertation, Universität Bonn.
- Kromrey, J. D. & Rendina-Gobioff, G. (2006). On knowing what we do not know – An empirical comparison of methods to detect publication bias in meta-analysis. *Educational and Psychological Measurement*, 66, 357-373.
- \*Ku, Y. L. & Ho, I. T. (2010). Dispositional factors predicting Chinese students' critical thinking performance. *Personality and Individual Differences*, 48, 54-58.
- \*Leikas, S., Mäkinen, S., Lönnqvist, J.-E. & Verkasalo, M. (2009). Cognitive ability x emotional stability interactions on adjustment. *European Journal of Personality*, 23, 329-342.

## Literatur

- \*LePine, J. (2003). Team adaptation and postchange performance: Effects of team composition in terms of members' cognitive ability and personality. *Journal of Applied Psychology*, 88, 27-39.
- \*LePine, J., Colquitt, J. A. & Erez, A. (2000). Adaptability to changing task contexts: Effects of general cognitive ability, conscientiousness, and openness to experience. *Personnel Psychology*, 53, 563-593.
- \*LePine, J., Hollenbeck, J. R., Ilgen, D. R. & Hedlund, J. (1997). Effects of individual differences on the performance of hierarchical decision-making teams: Much more than g. *Journal of Applied Psychology*, 82, 803-811.
- \*LePine, J. & Van Dyne, L. (2001). Voice and cooperative behavior as contrasting forms of contextual performance: Evidence of differential relationships with Big Five personality characteristics and cognitive ability. *Journal of Applied Psychology*, 86, 326-336.
- Leue, A. & Beauducel, A. (2008). A Meta-Analysis of reinforcement sensitivity theory: On performance parameters in reinforcement tasks. *Personality and Social Psychology Review*, 12, 353-369.
- Leue, A. & Lange, S. (2011). Reliability generalization: an examination of the positive affect and negative affect schedule. *Assessment*, 18, 487-501.
- Liepmann, D., Beauducel, A., Brocke, B. & Amthauer, R. (2008). *Intelligenz-Struktur-Test 2000 R* (2<sup>nd</sup>, erweiterte und überarbeitete ed.). Göttingen: Hogrefe.
- Light, R. J. & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.
- Lindenberger, U. & Baltes, P. B. (1997). Intellectual functioning in old and very old age: Cross-sectional results from the Berlin Aging Study. *Psychology and Aging*, 12, 410-432.
- \*Lounsbury, J. W., Sundstrom, E., Loveland, J. M. & Gibson, L. W. (2003). Intelligence, "Big Five" personality traits, and work drive as predictors of course grade. *Personality and Individual Differences*, 35, 1231-1239.
- \*Luciano, M., Leisser, R., Wright, M. J. & Martin, N. G. (2004). Personality, arousal theory and the relationship to cognitive ability as measured by inspection time and IQ. *Personality and Individual Differences*, 37, 1081-1089.
- \*Luciano, M., Wainwright, M. A., Wright, M. & Martin, N. G. (2006). The heritability of conscientiousness facets and their relationship to IQ and academic achievement. *Personality and Individual Differences*, 40, 1189-1199.
- Lutzenberger, W., Birbaumer, N., Flor, H., Rockstroh, B. & Elbert, T. (1992). Dimensional analysis of the human EEG and intelligence. *Neuroscience Letter*, 143, 10-14.



## Literatur

- \*Maurer, T. J., Lippstreu, M. & Judge, T. A. (2008). Structural model of employee involvement in skill development activity: The role of individual differences. *Journal of Vocational Behavior*, 72, 336-350.
- Mayer, K. U. & Baltes, P. B. (1996). *Die Berliner Altersstudie*. Berlin: Akademie-Verlag.
- \*McCrae, R. R. (1992). Openness to experience as a basic dimension of personality. *Imagination, Cognition, and Personality*, 13, 39-55.
- \*McCrae, R. R. & Costa, P. T. (1985). Updating Norman's "Adequate Taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology*, 49, 710-721.
- McCrae, R. R. & Costa, P. T. (1986). Clinical assessment can benefit from recent advances in personality psychology. *American Psychologist*, 41, 1001-1003.
- McCrae, R. R. & John, O. P. (1992). An introduction to the Five-Factor-Model and its applications. *Journal of Personality*, 60, 175-215.
- McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive gf-gc framework. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151-179). New York: Guilford Press.
- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2<sup>nd</sup> ed., pp.136-181). New York: Guilford Press.
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37, 1-10.
- \*McKenzie, J. & Tindell, G. (1993). Anxiety and academic achievement: Further furnace factor findings. *Personality and Individual Differences*, 15, 609-617.
- \*Miller, G. F. & Tal, I. R. (2007). Schizotypy versus openness and intelligence as predictors of creativity. *Schizophrenia Research*, 93, 317-324.
- \*Moses, L., Hall, C., Wuensch, K., De Urquidi, K., Kauffmann, P., Swart, W., Duncan, S. & Dixon, G. (2011). Are math readiness and personality predictive of first-year retention in engineering? *The Journal of Psychology*, 145, 229-245.
- Most, B. & Zeidner, M. (1995). Constructing personality and intelligence test instruments: Methods and Issues. In D. Saklofske & M. Zeidner (Eds.). *International handbook of personality and intelligence* (pp. 475-503). New York: Plenum.

## Literatur

- \*Motowidlo, S. J., Brownlee, A. L. & Schmit, M. J. (2008). Effects of personality characteristics on knowledge, skill, and performance in servicing retail customers. *International Journal of Selection and Assessment*, 16, 272-281.
- \*Mount, M. K., Barrick, M. R. & Strauss, J. P. (1999). The joint relationship of conscientiousness and ability with performance: Test of the interaction hypothesis. *Journal of Management*, 25, 707-721.
- \*Mount, M. K., Oh, I.-S. & Burns, M. (2008). Incremental validity of perceptual speed and accuracy over general mental ability. *Personnel Psychology*, 61, 113-139.
- \*Moutafi, J., Furnham, A. & Crump, J. (2003). Demographic and personality predictors of intelligence: A study using the Neo personality inventory and the Myers-Briggs Type Indicator. *European Journal of Personality*, 17, 79-94.
- \*Moutafi, J., Furnham, A. & Crump, J. (2006). What facets of openness and conscientiousness predict fluid intelligence score? *Learning and Individual Differences*, 16, 31-42.
- \*Moutafi, J., Furnham, A. & Patiel, L. (2004). Why is conscientiousness negatively correlated with intelligence? *Personality and Individual Differences*, 37, 1013-1022.
- \*Moutafi, J., Furnham, A. & Tsaousis, I. (2006). Is the relationship between intelligence and trait Neuroticism mediated by test anxiety? *Personality and Individual Differences*, 40, 587-597.
- National Research Council (1992). *Combining information: Statistical issues and opportunities for research*. Washington, DC: National Academy of Sciences Press.
- \*Neuman, G. A. & Wright, J. (1999). Team effectiveness: Beyond skills and cognitive ability. *Journal of Applied Psychology*, 84, 376-389.
- \*Newman, D. A. & Lyon, J. S. (2009). Recruitment efforts to reduce adverse impact: Targeted recruiting for personality, cognitive ability, and diversity. *Journal of Applied Psychology*, 94, 298-317.
- \*Newsome, S., Day, A. L. & Catano, V. M. (2000). Assessing the predictive validity of emotional intelligence. *Personality and Individual Differences*, 29, 1005-1016.
- \*Noftle, E. E. & Robins, R. W. (2007). Personality predictors of academic outcomes: Big Five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology*, 93, 116-130.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 66, 574-583.

## Literatur

- \*Oostrom, J. K., Born, M. P., Serlie, A. W. & van der Molen, H. T. (2011). A multimedia situational test with a constructed-response format. *Journal of Personnel Psychology, 10*, 78-88.
- Ostendorf, F. & Angleitner, A. (1994). The five-factor taxonomy: Robust dimensions of personality description. *Psychologica Belgica, 34*, 175-194.
- Ostendorf, F. & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae, Revidierte Fassung (NEO-PI-R)*. Göttingen: Hogrefe.
- Overton, R. C. (1998). A comparison of fixed effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods, 3*, 354-379.
- \*Parkinson, J. & Taggar, S. (2006). Intelligence, personality and performance on case studies. *Journal of Business and Psychology, 20*, 395-408.
- Peabody, D. (1987). Selecting representative trait adjectives. *Journal of Personality and Social Psychology, 52*, 59-71.
- \*Pearman, A. (2009). Basic cognition in adulthood: Combined effects of sex and personality. *Personality and Individual Differences, 47*, 357-362.
- Pervin, L. A. (1994). A critical analysis of current trait theory. *Psychological Inquiry, 5*, 552-567.
- \*Peterson, M. H., Griffith, R. L. & Converse, P. D. (2009). Examining the role of applicant faking in hiring decisions: Percentage of fakers hired and hiring discrepancies in single- and multiple-predictor Selection. *Journal of Business and Psychology, 24*, 373-386.
- \*Pincombe, J. L., Luciano, M., Martin, N. G. & Wright, M. J. (2007). Heritability of the NEO PI-R extraversion facets and their relationship with IQ. *Twin Researches and Human Genetics, 10*, 462-469.
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*, 322-338.
- \*Postlethwaite, B., Robbins, S., Rickerson, J. & McKinniss, T. (2009). The moderation of conscientiousness by cognitive ability when predicting workplace safety behavior. *Personality and Individual Differences, 47*, 711-716.
- Putnam, H. (1988). *Representation and reality*. Cambridge, MA: MIT Press.
- Raine, A., Reynolds, C., Venables, P. H. & Mednick, S. A. (2002). Stimulation seeking and intelligence: A prospective longitudinal study. *Journal of Personality and Social Psychology, 82*, 663-674.
- Raven, J., Raven, J. C. & Court, J. H. (2000). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Oxford: Oxford Psychologists Press.

## Literatur

- Rawlings, D. & Carnie, D. (1989). The interaction of EPQ extraversion with WAIS subtest performance under timed and untimed conditions. *Personality and Individual Differences*, 10, 453-458.
- Rawlings, D. & Skok, M. (1993). Extraversion, venturesomeness and intelligence in children. *Personality and Individual Differences*, 15, 389-396.
- Razoumnikova, O. (2003). Interaction of personality and intelligence factors in cortex activity modulation. *Personality and Individual Differences*, 35, 135-162.
- \*Reddock, C. M., Biderman, M. D. & Nguyen, N. T. (2011). The relationship of reliability and validity of personality tests to frame-of-reference instructions and within-person inconsistency. *International Journal of Selection and Assessment*, 19, 119-131.
- \*Reeve, C. L., Meyer, R. D. & Bonaccio, S. (2006). Intelligence-personality associations reconsidered: The importance of distinguishing between general and narrow dimensions of intelligence. *Intelligence*, 34, 387-402.
- \*Reichard, R. J., Riggio, R. E., Guerin, D. W., Oliver, P. H., Gottfried, A. W. & Gottfried, A. E. (2011). A longitudinal analysis of relationships between adolescent personality and intelligence with adult leader emergence and transformational leadership. *The leadership Quarterly*, 22, 471-481.
- \*Ridgell, S. D. & Lounsbury, J. W. (2004). Predicting academic success: general intelligence, "Big Five" personality traits, and work drive. *College Student Journal*, 38, 607-618.
- Rindermann, H. & Neubauer, A. C. (2001). The influence of personality on three aspects of cognitive performance: Processing speed, intelligence and school performance. *Personality and Individual Differences*, 30, 829-842.
- Rindfuss, R. R. (1991). "The young adult years: Diversity, structural change, and fertility." *Demography*, 28, 493-513.
- Roberts, B. W. & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, 126, 3-25.
- \*Roberts, M. J. (2002). The relationship between extraversion and ability. *Personality and Individual Differences*, 32, 517-522.
- \*Robie, C., Komar, S. & Brown, D. J. (2010). The effect of coaching and speeding on Big Five and Impression Management Scale scores. *Human Performance*, 23, 446-467.
- Robinson, D. L. (1985). How personality relates to intelligence test performance: Implications for a theory of intelligence, ageing research, and personality assessment. *Personality and Individual Differences*, 6, 203-216.

## Literatur

- Robinson, M. D. & Tamir, M. (2005). Neuroticism as mental noise: A relation between neuroticism and reaction time standard deviations. *Journal of Personality and Social Psychology, 89*, 107-114.
- \*Robson, S. M., Abraham, J. D. & Weiner, J. (2010). Characteristics of successful direct support professionals: An examination of personality and cognitive ability requirements. *International Journal of Selection and Assessment, 18*, 215- 219.
- \*Rode, J. C., Arthaud-Day, M. L., Mooney, C. H., Near, J. P. & Baldwin, T. T. (2008). Ability and personality predictors of salary, perceived career success in the initial career stage. *International Journal of Selection and Assessment, 16*, 292-299.
- Rodriguez, M. C. & Maeda, Y. (2006). Meta-analysis of coefficient alpha. *Psychological Methods, 11*, 306-322.
- Rolfhus, E. L. & Ackerman, P. L. (1999). Assessing individual differences in knowledge: Knowledge, intelligence, and related traits. *Journal of Educational Psychology, 91*, 511-526.
- \*Rosander, P., Bäckström, M. & Stenberg, G. (2011). Personality traits and general intelligence as predictors of academic performance: A structural equation modelling approach. *Learning and Individual Differences, 21*, 590-596.
- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin, 86*, 638-641.
- Rosnow, R. L. & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44*, 1276-1284.
- Rothstein, H. R. & Bushman, B. J. (2012). Publication bias in psychological science: Comment on Ferguson and Brannick (2012). *Psychological Methods, 17*, 129-136.
- Rouse, S. V. (2007). Using reliability generalization methods to explore measurement error: An illustration using the MMPI-2 PSY-5 scales. *Journal of Personality Assessment, 88*, 264-275.
- \*Saggino, A. & Balsamo, M. (2003). Relationship between WAIS-R intelligence and the Five-Factor Model of personality in a normal elderly sample. *Psychological Reports, 92*, 1151-1161.
- Saklofske, D. H. & Kostura, D. D. (1990). Extraversion–introversion and intelligence. *Personality and Individual Differences, 11*, 547-551.
- Saklofske, D. H. & Zeidner, M. (1995). *International Handbook of Personality and Intelligence*. New York: Plenum Press.

## Literatur

- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., De Fruyt, F. & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European Community. *Journal of Applied Psychology*, 88, 1068-1081.
- Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, 30, 507-514.
- Scarr, S. (1985). Constructing Psychology: Making facts and fables for our times. *American Psychologist*, 40, 499-512.
- Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist*, 49, 304-313.
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Schmidt, F. L. & Le, H. (2004). Software for the Hunter-Schmidt meta-analysis methods. University of Iowa, Department of Management & Organization, Iowa City, IA, 42242.
- Schuler, H. (2002). Emotionale Intelligenz – ein irreführender und unnötiger Begriff. *Zeitschrift für Personalpsychologie*, 1, 138-140.
- \*Schulte, M. J., Ree, M. J. & Carretta, T. R. (2004). Emotional intelligence: Not much more than g and personality. *Personality and Individual Differences*, 37, 1059-1068.
- Schulze, R & Roberts, R. D. (2006). Assessing the Big Five. Development and validation of the openness conscientiousness extraversion agreeableness neuroticism index condensed (OCEANIC). *Zeitschrift für Psychologie*, 214, 133-149.
- \*Schurr, K. T., Ruble, V. E. & Henriksen, L. W. (1988). Relationships of Meyers-Briggs Type Indicator personality characteristics and self-reported academic problems and skill ratings with Scholastic Aptitude Test scores. *Educational and Psychological Measurement*, 48, 187-196.
- \*Schweizer, K. (2002). Does impulsivity influence performance in reasoning? *Personality and Individual Differences*, 33, 1031-1043.
- \*Sharp, E. S. & Gatz, M. (2010). Cognitive engagement and cognitive aging: Is openness protective? *Psychology and Aging*, 25, 60-73.
- \*Sharp, E. S., Reynolds, C. A., Pedersen, N. L. & Gatz, M. (2010). Cognitive engagement and cognitive aging: Is openness protective? *Psychology and Aging*, 25, 60-73.
- \*Silvia, P. J. & Sanders, C. E. (2010). Why are smart people curious? Fluid intelligence, openness to experience, and interest. *Learning and Individual Differences*, 20, 242-245.

## Literatur

- \*Simon, A., Fontana, D. & Williams, E. (1986). Performance on tests of study habits, neuroticism, extraversion, academic self-concept, intelligence and A level by polytechnic students as focusers and scanners on a concept matrix test. *Personality and Individual Differences*, 7, 229-231.
- Spearman, C. (1904). "General Intelligence" objectively and determined and measured. *The American Journal of Psychology*, 15, 201-292.
- \*Spector, P. E., Schneider, J. R., Vance, C. A. & Hezlett, S. A. (2000). The relation of cognitive ability and personality traits to assesement center performance. *Journal of Applied Social Psychology*, 30, 1474-1491.
- \*Sperbeck, D. J., Whitbourne, S. K. & Hoyer, W. J. (1986). Age and openness to experience in autobiographical memory. *Experimental Aging Research*, 12, 169-172.
- Spinath, F. M. & Johnson, W. (2011). Behavior Genetics. In T. Chamorro-Premuzic, S. Stumm & A. Furnham (Eds.), *The Wiley-Blackwell Handbook of Individual Differences*, (pp. 271-304). West Sussex: Wiley-Blackwell.
- Stankov, L. (2000). Complexity, metacognition, and fluid intelligence. *Intelligence*, 28, 121-143.
- Stankov, L., Boyle, G. L. & Cattell, R. B. (1995). Models and paradigms in personality and intelligence research. In D. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence* (pp. 15-43). New York: Plenum.
- \*Staudinger, U. M., Maciel, A. G., Smith, J. & Baltes, P. B. (1998). What predicts wisdom-related performance? A first look at personality, intelligence, and facilitative experimental contexts. *European Journal of Personality*, 12, 1-17.
- \*Steinmayr, R., Dinger, F. C. & Spinath, B. (2010). Parents' education and children's achievement: The role of personality. *European Journal of Personality*, 24, 535-550.
- \*Stough, C., Brebner, J., Nettelbeck, T., Cooper, C. J., Bates, T. & Mangan, G. L. (1996). The relationship between intelligence, personality and inspection time. *British Journal of Psychology*, 87, 255-268.
- Stumm, S., Chamorro-Premuzic, T & Ackerman, P. L. (2011). Re-visiting intelligence-personality associations: Vindicating intellectual invetsment. In T. Chamorro-Premuzic, S. Stumm & A. Furnham (Eds.), *The Wiley-Blackwell Handbook of Individual Differences*, (pp. 217-241). West Sussex: Wiley-Blackwell.
- Süß, H.-M. (2001). Prädiktive Validität der Intelligenz im schulischen und außerschulischen Bereich. In E. Stern & J. Guthke (Eds.), *Perspektiven der Intelligenzforschung*. Lengerich: Pabst Science Publisher, 109-135.

## Literatur

- Süß, H.-M. & Beauducel, A. (2011). Intelligenztests und ihre Bezüge zu Intelligenztheorien. In L. F. Hornke, M. Amelang & M. Kersting (Eds.), *Enzyklopädie der Psychologie: Leistungs-, Intelligenz- und Verhaltensdiagnostik 3* (pp. 15-43). Göttingen: Hogrefe.
- \*Taggar, S., Hackett, R. & Saha, S. (1999). Leadership emergence in autonomous work teams: Antecedents and outcomes. *Personnel Psychology*, *52*, 899-926.
- \*Te Nijenhuis, J. & van der Flier, H. (2000). Differential prediction of immigrant versus majority group training performance using cognitive ability and personality measures. *International Journal of Selection and Assessment*, *8*, 54-60.
- Tellegen, A. (1982). *Brief manual for the Multidimensional Personality Questionnaire (MPQ)*. Minneapolis, MN: Author.
- Terracciano, A., McCrae, R.R. & Costa, P.T. Jr. (2010). Intra-individual change in personality stability and age. *Journal of Research in Personality*, *44*, 31-37
- Thatcher, R. W., North, D. & Biver, C. (2008). Intelligence and EEG phase reset: A two compartmental model of phase shift and lock. *NeuroImage*, *42*, 1639-1653.
- \*Thompson, R. F. & Perlini, A. H. (1998). Feedback and self-efficacy, arousal, and performance of introverts and extraverts. *Psychological Reports*, *82*, 707-716.
- Thorndike, R. L. (1982). *Applied Psychometrics*. Boston: Houghton Mifflin.
- \*Tracey, J. B., Sturman, M. C. & Tews, M. J. (2007). Ability versus personality, factors that predict employee job performance. *Cornell Hotel and Restaurant Administration Quarterly*, *48*, 313-322.
- Tupes, E. C. & Christal, R. E. (1958). *Stability of personality trait rating factors obtained under diverse conditions* (USAF WADC Tech. Note No. 58-61). Lackland Air Force Base, TX: U.S. Air Force.
- Tupes, E. C. & Christal, R. E. (1961). *Recurrent personality factors based on trait ratings* (USAF ASD Tech. Rep. No. 61-97). Lackland Air Force Base, TX: U.S. Air Force.
- \*Uttl, B. & Kibreab, M. (2011). Self-report measures of prospective memory are reliable but not valid. *Canadian Journal of Experimental Psychology*, *65*, 57-68.
- Vacha-Haase, T., Henson, R. & Caruso, J. C. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement*, *62*, 562-569.
- Vernon, P. A. (1987). *Speed of information processing and intelligence*. Norwood, NJ: Ablex.
- \*Wainwright, M. A., Wright, M. J., Luciano, M., Geffen, G. M. & Martin, N. G. (2008). Genetic covariation among facets of openness to experience and general cognitive ability. *Twin Research and Human Genetics*, *11*, 275-286.



## Literatur

- \*Warwick, J. & Nettelbeck, T. (2004). Emotional intelligence is...? *Personality and Individual Differences*, 37, 1091-1100.
- Watson, G. & Glaser, E. M. (1964). *Watson-Glaser Critical Thinking Appraisal Manual: Forms YM and ZM*. New York: Harcourt, Brace, and World.
- Wechsler, D. (1981) *Manual for the Wechsler Adult Intelligence Scale—Revised*. Psychological Corporation, New York.
- Westmeyer, H. (1995). Persönlichkeitspsychologie zwischen Realismus und Konstruktivismus. In K. Pawlik (Ed.), *Bericht über den 39. Kongreß der Deutschen Gesellschaft für Psychologie 1994 in Hamburg*. Göttingen: Hogrefe, 109-135.
- Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals in meta-analysis. *Journal of Applied Psychology*, 75, 315-321.
- \*Wickett, J. C. & Vernon, P. A. (2000). Replicating the movement time-extraversion link... with a little help from IQ. *Personality and Individual Differences*, 28, 205-215.
- \*Williams, K. M., Nathanson, C. & Paulhus, D. L. (2010). Identifying and profiling scholastic cheaters: Their personality, cognitive ability, and motivation. *Journal of Experimental Psychology*, 16, 293-307.
- Wittmann, W. W. (1985). *Evaluationsforschung*. Heidelberg: Springer.
- Wolf, M. B. & Ackerman, P. L. (2005). Extraversion and intelligence: A meta-analytic investigation. *Personality and Individual Differences*, 39, 531-542.
- Wonderlic, E. (1992). *Wonderlic Personnel Test*. IL: Libertyville.
- \*Yeo, G. & Neal, A. (2008). Subjective cognitive effort: A model of states, traits, and time. *Journal of Applied Psychology*, 93, 617-631.
- Yerkes, R. M. & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482.
- Youngstrom, E. A. & Green, K. W. (2003). Reliability generalization of self-report of emotions when using the Differential Emotions Scale. *Educational & Psychological Measurement*, 63, 279-295.
- Zeidner, M. (1995). Personality trait correlates of intelligence. In D. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence* (pp. 299-319). New York: Plenum.
- Zeidner, M. & Matthews, G. (2000). Intelligence and personality. In R. Sternberg (Ed.), *Handbook of intelligence*. New York, NY, US: Cambridge University Press.

## *Literatur*

- Ziegler, M., Danay, E., Heene, M., Asendorpf, J. & Bühner, M. (2012). Openness, fluid intelligence, and crystallized intelligence: Toward an integrative model. *Journal of Research in Personality, 46*, 173-183.
- \*Ziegler, M., Schmukle, S., Egloff, B. & Bühner, M. (2010). Investigating measures of achievement motivation(s). *Journal of Individual Differences, 31*, 15-21.
- \*Zimprich, D., Allemand, M. & Dellenbach, M. (2009). Openness to experience, fluid intelligence, and crystallized intelligence in middle-aged and old adults. *Journal of Research in Personality, 43*, 444-454.

## Anhang

### Anhang

Bevor der Schätzer  $R_0$  berechnet werden kann, müssen die Absolutbeträge ( $r_{\text{Abs}}$ ) gebildet werden, in dem der Mittelwert der berichteten Korrelationen ( $\bar{r}$ ) von der jeweiligen Korrelation ( $r_i$ ) subtrahiert wird.

$$r_{\text{Abs}} = (| r_i - \bar{r} |). \quad (\text{A1})$$

Im Anschluss werden die Ränge für  $r_{\text{Abs}}$  gebildet. Wenn der Mittelwert der Korrelation ( $\bar{r}$ ) größer ist als die jeweilige Korrelation ( $r_i$ ), bekommt der gebildete Rang ein negatives Vorzeichen. Als nächstes wird  $\gamma$  bestimmt mit

$$\gamma = k + n_g. \quad (\text{A2})$$

Wobei  $k$  die Anzahl der inkludierten Studien ist und  $n_g$  der größte negative Rang. Der berechnete Wert von  $\gamma$  wird danach in die Formel zur Bestimmung von  $R_0$  eingesetzt.

$$R_0 = \gamma - 1. \quad (\text{A3})$$

In der Literatur gibt es unterschiedliche Vorschläge wie weiter zu verfahren ist. In Duval (2005) wird eine substantielle Asymmetrie, die adjustiert werden sollte, ab einem Wert von  $R_0 = 1$  angenommen. Hingegen wird in Kromrey und Rendina-Gobioff (2006) ein Bias ab einem Wert von  $R_0 = 3$  adjustiert. Die Originalarbeiten von Duval und Tweedie (2000a, 2000b) sowie Duval (2005) berücksichtigend, wurde in der vorliegenden Metaanalyse eine Adjustierung ab  $R_0 = 1$  durchgeführt.

Das zweite Symmetriemaß  $L_0$  basiert auf der Summe aller Ränge, die zu positiven Korrelationen gehören ( $S_{\text{rank}}$ ).  $S_{\text{rank}}$  wird zur Anzahl aller Korrelationen ( $k$ ) in Beziehung gesetzt.

$$L_0 = \frac{4S_{\text{rank}} - k(k+1)}{2k-1}. \quad (\text{A4})$$

Eine Adjustierung der Asymmetrie soll nach Duval (2005) ab einem Wert von  $L_0 = 1$  erfolgen.

## **Erklärung**

Ich versichere, dass ich diese Dissertationsschrift selbst und ohne andere als die darin angegebenen Hilfsmittel angefertigt sowie die wörtlich oder inhaltlich übernommenen Stellen als solche gekennzeichnet habe.

Bonn, März 2013